AN INTRODUCTION TO THEORY AND PRACTICE OF PSYCHOLOGY

MACMILLAN AND CO., LIMITED
LONDON · BOMBAY · CALCUTTA · MADRAS
MELBOURNE

THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO
DALLAS · ATLANTA · SAN FRANCISCO

THE MACMILLAN COMPANY
OF CANADA LIMITED
TORONTO

AN INTRODUCTION TO THEORY AND PRACTICE OF PSYCHOLOGY

BY

LL. WYNN JONES

M.A. (OXON), PH.D. (LEIPZIG)

EXAMINER IN PSYCHOLOGY TO THE UNIVERSITY OF LONDON

MACMILLAN AND CO., LIMITED
ST. MARTIN'S STREET, LONDON
1934

PRINTED IN GREAT BRITAIN
BY R. & R. CLARK, LIMITED, EDINBURGH

PREFACE

In the twentieth century there has occurred two related psychological events of supreme importance. On the one hand, the practice of the science of psychology for the first time vitally affected the social and economic life of the people. On the other hand, there is the attempt by *psychologists* to base their science solely on *psychological* principles. This allows ordinary mortals to advance into realms which were uncharted even for masterminds possessing the highest scientific credentials—Aristotle, Descartes, Leibniz, Wundt.

Besides, a mixture of psychological, physiological, physical and metaphysical "principles" in a single book would certainly confuse the beginner. In this book, then, it will be assumed that psychology is an independent science based on fundamental principles which can be applied in such widely different fields as education, industry and medicine. There follows the corollary that whenever there is an appeal against a psychological finding the court must necessarily be psychological as, from the nature of the case there is not, and cannot be, an external court of appeal.

It is now agreed that "experimental psychology" without theory is futile and that "general psychology" without experiment is sterile. It is difficult to realise that Dalton, Davy, Faraday, Darwin and Joule received no training in experimental methods, and that even Clerk Maxwell, Stokes and Kelvin had no opportunity to do experimental work while at college.[1] But revolutionary as the advance of physical science has been in the nineteenth century, it may well be eclipsed by that of psychology in the twentieth.

Let Dean Seashore[2] speak of his own psychological laboratory at the State University of Iowa: (In 1900) "I had no assistants.

[1] D. M. Turner, *History of Science Teaching in England*, p. 122. Cf. also *Nature*, January 28, 1928.

[2] *History of Psychology in Autobiography*, edited by C. Murchison, vol. 1, pp. 261 and 264.

I was my own stenographer, my own mechanician, and my own bottle-washer". (In 1930) "We are moving into new quarters which will furnish perhaps the most generous allowance of space and accessories for a department of psychology anywhere. The various branches of applied psychology such as clinical psychology, speech pathology, psychology of the child and child welfare, the psychology of art, and the psychology of music are assembled in one building around the central unit. There are about seventy-five rooms available for the graduate work in psychology, and some thirty-five additional for co-operating services."

The would-be specialist in psychology has usually taken Latin and French at school. He will now find four ancillary subjects desirable: philosophy, physiology, statistics and German. But as the doctrine that "Arts is Arts and Sci. is Sci. and never the twain shall meet" is not yet defunct, he may be fettered unless he is determined at any cost to follow a congruent and cultural course. The unsophisticated who believes in the logical coherency of labels may be informed that at some seats of learning he may specialise in psychology for the B.A. degree, at others for the B.Sc., at some for either, and at others for neither! But those who are acquainted with recent changes in this and in other countries are aware that, in framing curricula, freedom without licence is rapidly becoming an attainable ideal.

Then there is the student of educational psychology. I once saw a report advocating the abolition of psychology as a subject for teachers in training on the ground that it is too difficult for their immature minds! Possibly so for a minority who unfortunately seem to miss their vocation, but for the rest surely an essential study demanding sufficient *time* and *space* and *trained instructors*.

Portions of this book will be familiar to those who attended my lectures at the Universities of Missouri, Iowa, Minnesota, Michigan, Wisconsin and Pittsburgh in 1929. Many of the experiments can be taken with children as well as with adults and have been extensively used with graduate students taking the Diploma in Education in their fourth year. Portions of the

book have been used in classes in the Faculty of Medicine in preparation for the Diploma in Psychological Medicine, also in various classes in industrial psychology.

Topics which were considered too advanced for an introductory course have been deliberately omitted.

The book is based on a system which has exercised a paramount influence on psychological progress. It is known as the Spearman school from the name of its founder, as the Noegenetic school from the nature of its principles, and as the Factor school from the results of its analyses. Most of the experiments selected can be performed without expensive apparatus, and the simpler statistical methods of dealing with psychological data have been incorporated so as to combine in one volume an introduction to both the theory and practice of psychology.

I am indebted to Professor G. M. Whipple for permission to incorporate the directions for using the Cornell Form Board as given in his *Manual of Mental and Physical Tests* (Messrs. Warwick and York, Baltimore).

I am also indebted to Dean C. E. Seashore for permission to incorporate his directions for using his Measures of Musical Talent as given in his *Manual of Instructions and Interpretations for Measures of Musical Talent* (Columbia Graphophone Company, New York City).

For smoothing many an uncouth sentence I am indebted to my wife—sometime holder of the Senior Prize in English Literature at University College, London.

In compiling the book I have been influenced by many authors, but for its plan and all imperfections I accept sole responsibility. In conclusion I must record that I cannot adequately express my obligations to Professor Spearman, not only for his epoch-making contributions to psychology but also for that unique inspiration which remains the permanent possession of those who have served on his staff.

LL. WYNN JONES

UNIVERSITY OF LEEDS
September 1934

CONTENTS

CHAPTER I

INTRODUCTORY

> When you leave unconscious matter, and come to deal with con scious minds, a wholly new technique has to be adopted.
>
> The late Right Hon. THE EARL OF BALFOUR
> *Journ. of the Nat. Inst. of Indust. Psychology*
> vol. 2, No. 3, July 1924

IN every science it is very desirable but also very difficult to employ only terms explicit in meaning. Especially so in psychology, where the student has to study the processes of his own mind as well as of the minds of others. If, on the one hand, he attempts to discard psychological terms to any radical degree he will soon find himself in the quicksands of loose thinking. But on the other hand, his advance may also be effectively checked by the bewildering multiplicity of psychological terms, especially when he finds considerable disagreement amongst psychologists as to their meaning. It is true that the psychologist, like other scientists, must coin a new word if progress demands it, but the dangers of loose verbiage were realised at the time of Aristotle.

The following passage written two thousand years later well expresses it:

> Although we think we govern our words, and prescribe it well *loquendum ut vulgus sentiendum ut sapientes*; yet certain it is that words, as a Tartar's bow, do shoot back upon the understanding of the wisest, and mightily entangle and pervert the judgment. So is it almost necessary, in all controversies and disputations, to imitate the wisdom of the mathematicians, in setting down in the very beginning the definitions of our words and terms, that others may know how we accept and understand them, and whether they concur with us or no. For it cometh to pass, for want of this, that we are sure to end there where we ought to have begun, which is, in questions and differences about words.[1]

[1] F. Bacon, *The Advancement of Learning*, Book 2, xiv. 11.

Another obstacle is that many textbooks include a physiological account of the senses and of the nervous system, either in the introductory chapters or as an appendix. Now, as McDougall has pointed out, all this appears to be of very secondary importance for the beginner. Moreover, "our knowledge of the functions of the nervous system is very rudimentary, and as regards many of those of greatest interest to psychology we are still entirely in the dark".[1] Equally emphatic is the weighty evidence of Lashley: "Psychology is today a more fundamental science than neurophysiology. By this I mean that the latter offers few principles from which we may predict or define the normal organisation of behaviour, whereas the study of psychological processes furnishes a mass of factual material to which the laws of nervous action in behaviour must conform."[2]

More confusing still is the intermingling of physiological and psychological terms and theories which may leave the reader uncertain, if not bewildered, concerning the bases of his science. It is true that some knowledge of physiology is essential for the psychologist, and obviously this could best be imparted by a physiologist, preferably by direct teaching or, failing that, through the medium of one of the many excellent physiological textbooks. On these grounds a presentation in simple psychological terms may have distinct advantages, especially in the case of students whose time is limited.

Now it is a striking fact that there are some fifty common and well-established verbs in English without which the language would be paralysed as a medium of thought. I apprehend, perceive, see, hear, taste, smell, touch, am hot, am cold, have pain, hunger, thirst, experience pressure, image, remember, forget, feel pleasure (or unpleasure), am excited, am tranquil, am tense, am relaxed, am angry, fear, am curious, am disgusted, am bored, am tired, hope, wonder, like, love, hate, will, move, strive, control, repress, desire, expect, think, imagine, reason, judge, compare, discriminate, imitate, suggest, sympathise with, believe, doubt . . .

[1] W. McDougall, *An Outline of Psychology*, Preface, p. xi, 1924.

[2] K. S. Lashley, "Basic Neural Mechanisms in Behaviour", *Psychol. Rev.* vol. 37, No. 1, 1930, p. 24.

Such words in their appropriate spatial and temporal settings provide the means by which we describe our experiences. It would be rash to assume that all educated persons interpret these words alike, for, as McDougall[1] points out, language remains an inadequate instrument for describing our experiences, in spite of all the efforts of literary men and of psychologists to render it more precise, But it will be assumed that the meanings of the above words are definite enough and universal enough to enable them to be profitably used when elementary accounts of mental experiences are given. It is needless to add that the use of some of the above words, as in the following phrases, does not here concern us: I *see* your point of view. I *touch* the fringe of the subject. Your argument leaves me *cold.* The news *pains* me.

Up to this point no basic principles for future guidance have been formulated, To prepare the ground for such formulation it is advisable to inquire what meanings should be attached to three words: Experience, Intuition, Belief. Each of these words has had a long history, but here is hardly the place to do more than indicate and accept the meanings given by Spearman in his attempt to provide psychology with a solid foundation.

EXPERIENCE.—The first word on the list—I apprehend—deserves a special explanation on account of its importance. One cannot apprehend without "experience". It is necessary, therefore, to inquire what "experience" is intended to signify. It may mean knowledge derived from proof furnished by one's own senses, or, it may mean something lived, undergone, enjoyed, or the like.[2] What sort of knowledge is really furnished by one's own senses is seen, however, on reflection to be open to serious doubt, and weighty reasons are given by Spearman why the second meaning should be adopted, that is to say, "experience" is something lived, undergone, enjoyed, and the like.

INTUITION.—In common parlance this word is taken to

[1] W. McDougall, *op. cit.* p. 4.

[2] C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition*, 1923, p. 36.

apply to a kind of implicit apperception on grounds which cannot be analysed. McDougall[1] gives this meaning his blessing. "Some minds, even of persons who have little power of abstract thinking and little command of abstract terms, work in this way with marvellous subtlety; they are properly said to be gifted by intuition. Some women possess it in very high degree; young children, whose command of language is very slight, may exhibit it; and even in the higher animals, especially the dog, it is not altogether lacking." On the other hand, Valentine did not find women to be superior in their intuitive judgment of character qualities.[2]

It is fortunately not necessary at this point to accept or to deny this widely held view. For our present purpose "intuitive" has another meaning. It is taken to mean knowledge that is self-evident and unmediated.[3] And in this sense the three laws shortly to be enunciated show whereby belief is derived from intuitive evidence.

Belief.—The ancient meaning of belief was essentially an acceptance of anything as true without adequate rational grounds. Sometimes it is taken to mean an acceptance on the ground of testimony or of reason. Spearman, however, employs the word to denote quite generally the acceptance of anything as true, irrespective of the particular nature of the grounds for doing so. Nevertheless the immediate scope of Spearman's psychological principles includes only such beliefs as rest upon adequate evidence.[4]

We may now proceed to state and illustrate three qualitative principles or ultimate psychological laws which have been formulated by Spearman:

FIRST PRINCIPLE: THE APPREHENSION OF EXPERIENCE

Any lived experience tends to evoke immediately a knowing of its direct attributes and its experiencer.[5]

[1] W. McDougall, *op. cit.* p. 391.

[2] C. W. Valentine, "The Relative Reliability of Men and Women in Intuitive Judgments of Character", *Brit. Journ. of Psychol.* vol. 19, 1929.

[3] C. Spearman, *The Abilities of Man*, 1927, p. 227, footnote.

[4] C. Spearman, *The Nature of Intelligence*, p. 57.

[5] *Ibid.* p. 48.

Illustration.—Let a tuning fork be sounded. The hearer immediately becomes aware of, or apprehends, certain attributes of the experience: (1) It has a certain quality, it is a tone and not a noise. (2) It has a certain intensity which gradually diminishes. (3) It has a definite spatial character so that the hearer has a more or less definite awareness of its location. (4) It has a definite temporal character, that is, it begins and ends at definite times. Moreover, the hearer may derive pleasure from the experience and wish it to continue. In addition, the hearer possesses the idea of himself as being the experiencer.

SECOND PRINCIPLE: THE EDUCTION OF RELATIONS

The presenting of any two or more characters tends to evoke immediately a knowing of relation between them.[1]

Illustration.—The sounding of two tuning forks of frequency C and G respectively tends to evoke a knowing of some relation between them, *e.g.* "one is higher in pitch than the other", or "I prefer the G sound", or "the interval is a fifth" and the like, according to the attitude of the experiencer. The verbal expression of the relation evoked will, of course, depend on various factors such as the language employed by the experiencer, his musical training and his introspective skill.

THIRD PRINCIPLE: THE EDUCTION OF CORRELATES

The presenting of any character together with a relation tends to evoke immediately a knowing of the correlative character.[2]

Illustration.—Let the character presented be the sound of the F fork and let the relation presented be that evoked from sounding the C and G forks, then the correlative character which tends to be evoked is the fifth built on F as fundamental.

A simpler illustration in non-musical terms would be:

> Character presented: arm, relation presented: leg—foot, correlative character evoked: hand.

[1] C. Spearman, *The Nature of Intelligence*, p. 63.
[2] *Ibid.* p. 91.

These three principles are couched in psychological terms and explanations of physiological order are not involved. Such indeed is one criterion which basic psychological principles should satisfy. Moreover, explanations of physiological order are unknown in this connection, and are likely to remain so.

These three principles also represent a necessary order of sequence. First the apprehension of experience, then the eduction of relations which precedes the eduction of correlates.

It is obvious that in all teaching a consideration of this sequence is very important. In fact, the systematic application of these principles in the field of education is an urgent problem which has not even yet been sufficiently investigated.

Finally, to psychology founded on such a basis "falls the exclusive right and duty of settling what respective mental operations actually occur when any person entertains a belief that is either erroneous or its converse truthful, using these terms in the most ordinary sense 'without prejudice' ".[1]

In the experiments which follow, practice will be obtained (1) in giving an exact account of experiences, *i.e.* an introspective report, (2) in the study of the nature and magnitude of the mental differences between individuals.

THE NATURE OF INTROSPECTION.—The attitude of a writer towards the method of introspection tells a good deal about his psychology. Wundt, the founder of experimental psychology, would not allow the testing in his laboratory of individuals who could not systematically scrutinise their mental processes, that is, introspect. On the other hand, one of his first pupils, J. McKeen Cattell,[2] wrote in 1904: "It seems to me that most of the research work that has been done by me or in my laboratory is nearly as independent of introspection as work in physics or in zoology". In 1911 C. S. Myers[3] wrote: "I want to protest as strongly as I can against the notion that any useful purpose can be served, so far as psychology is concerned, by collecting

[1] C. Spearman, "The Origin of Error", *Journ. of General Psychol.* vol. 1, No. 1, 1928, p. 31.

[2] J. McKeen Cattell, *Psychology in America*, p. 32.

[3] C. S. Myers, "The Pitfalls of 'Mental Tests'", *Brit. Med. Journ.*, January 28, 1911.

masses of psychological data with the help of an army of untrained observers. . . . To neglect introspection in psychological experiment is usually to court certain disaster."

In 1912 Wertheimer founded a school of Gestalt psychology.[1] The leaders Wertheimer, Koffka and Köhler protested against the analysis into elements like sensations, images and feelings and substituted for them the concept of Gestalten or organised wholes. They held that "mental content is not a mosaic of solid unalterable things",[2] as would appear to be the case from the writings of many of the earlier psychologists. They also held that "much of current introspection seems to be rather sterile and, in an odd contrast to its ambitions, to lead research away from the more urgent problems".[3]

In 1913 Watson made a still more drastic break with orthodox psychology and founded behaviourism, and sought to abolish not only introspection but also consciousness as they were considered to be terms which could only hamper progress. Watson therefore confined his researches to the objective behaviour of men and animals. This is the so-called "psychology of stimulus and response". Watson has carried out important researches on problems of objective behaviour, but he soon found it necessary to include verbal report among his explicit responses.

Köhler, while attacking the extreme introspectionist's search for a "pure sensory process"—a will-o'-the-wisp which is not hunted by the modern psychologist—nevertheless advocates introspection in the sense of a phenomenological description of consciousness. Cattell's attitude may become clearer if it is remembered that for the last fifty years his major task has been the investigation of individual differences—a science which was largely his own creation. Long before behaviourism was known as a movement Cattell had consistently paid more regard to the objective responses of individuals than to the refine-

[1] The advanced student should compare the views of this school with those of earlier writers such as von Ehrenfels, Husserl, Cornelius, and Meinong.

[2] K. Koffka, "Introspection and the Method of Psychology", *Brit. Journ. of Psychol.* vol. 15, Part 2, October 1924, p. 151.

[3] W. Köhler, *Gestalt Psychology*, 1929, p. 19.

ments of Wundtian introspection. Moreover, an investigator with Cattell's unrivalled experience would avoid pitfalls which would spell disaster to the average investigator who decided to throw introspection overboard. On the other hand Wundt, before the end of his academic career, gave his blessing to experimental pedagogy, and for certain experiences children can undoubtedly give reliable introspections;[1] more reliable perhaps in one respect than those of sophisticated adults, who are liable to get out of an experiment what has already been put in it as a preconceived theory.

Critics of the method of introspection are fond of harping on one or more of the following arguments: (1) Introspection is always actually retrospection. (2) The act of introspection cannot fail to disturb the system under observation as it is a process *in* this system. (3) Introspection is impossible as the observer and the observed are one. (4) The act of attention in introspection can only extend to the inner experiences of the introspecting individual and there can be no "science" of the individual. In other words, introspection is too subjective, thus when a subject A reports "red" he may mean a different quality from that experienced by subject B who similarly reports "red".

Weighty answers to these contentions have been summarised by Dawes Hicks.[2] It would be difficult to name a more competent authority. The upshot would seem to be (1) that if some of these arguments against introspection were pressed, they would prove equally fatal to external observation and consequently to science of any kind, and (2) the difficulties of introspecting have been exaggerated so that the contentions of the critics hardly warrant the scrapping of the method of introspection but rather its refinement. It is, at least, usually possible to detect abnormalities such as colour blindness. Introspection, then, although a dangerous weapon, would seem to be indispensable.

[1] I employed psychophysical methods successfully with children aged five. Cf. *Pädogogische-psychologische Arbeiten*, vol. 2, Leipzig, 1911.

[2] G. Dawes Hicks, "The Nature of Introspection", *Aristotelian Soc. Suppl.* vol. 7, 1927.

THE NATURE AND MAGNITUDE OF MENTAL DIFFERENCES

"The differences between one man and another are always (we shall find) a matter of 'more or less'—seldom, if ever, a question of presence or absence, of 'all or none'.

> 'Virtuous and vicious ev'ry man must be
> Few in th' extreme, but all in a degree.'

There are, in fact, no such things as mental types; there are only mental tendencies."[1]

That this is true of a physical trait like height is obvious, and the following table gives the frequency-distributions of statures of a sample of adult males born in Ireland as measure by the Anthropometric Committee of the British Association:

Height in Inches.	Frequency F.	Cumulative F.
73→	3	345
72→	10	342
71→	15	332
70→	25	317
69→	40	292
68→	62	252
67→	73	190
66→	58	117
65→	33	59
64→	15	26
63→	7	11
62→	2	4
61→	2	2
	345	

If it be assumed that each individual may be measured to the nearest quarter of an inch, then it is conventional to include all of heights 73 inches, 73·25 inches, 73·5 inches or 73·75 inches in the compartment indicated by 73 in the first column of the table. From the second column it is seen they number 3.

[1] C. Burt, Presidential Address to the Psychol. Section, *Brit. Assoc. Report*, 1923.

From the first two columns it will be seen that the number of adults whose height was 61 inches or over but less than 62 was 2. From the third column it will be seen that the number of adults whose height is less than 66 inches is 59. The most common height met is evidently between 67 and 68 inches.

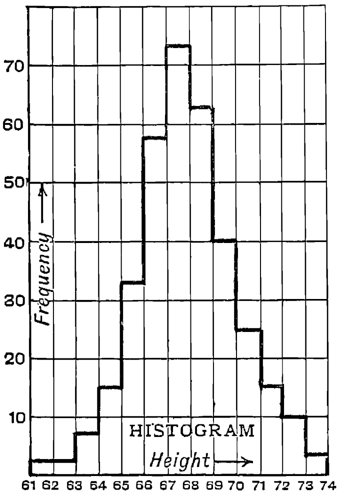

FIG. 1.

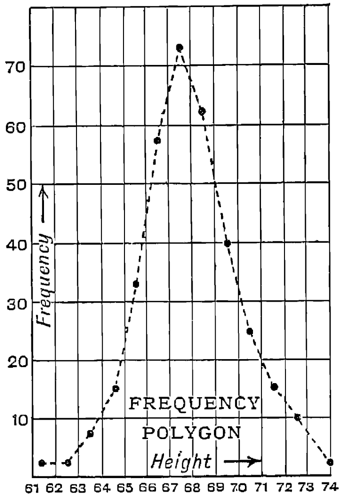

FIG. 2.

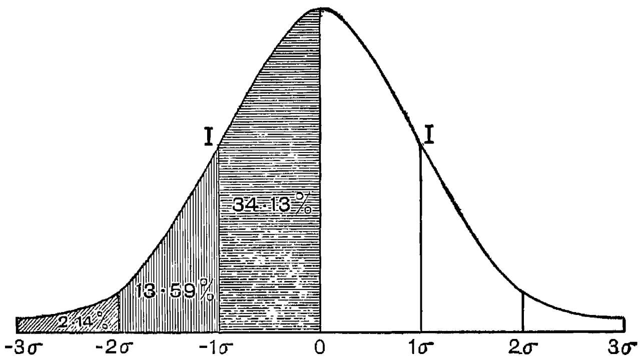

FIG. 3.

The frequency-distribution may be represented graphically either as a *histogram* or as a *frequency polygon*. To plot a histo-

gram draw a rectangle on each interval of height, and let the height of the rectangle be equal to the number of adults within the interval. To plot a frequency polygon erect an imaginary vertical line at the mid-point of each interval and let its height be equal to the number of men within the interval. Place a dot at the end of each imaginary line. The frequency polygon is formed by joining these dots. It is obvious that if the intervals, instead of being an inch, were made smaller and smaller, and at the same time the frequencies be increased by taking bigger samples, then both the histogram and the frequency polygon will approach more and more to a smooth curve which is called a *frequency curve.*

Fisher[1] points out that where the variation is continuous the frequency diagram should be in the form of a histogram referring to the actual sample of observations, for one cannot assume the resulting frequency polygon to be an approximation to the continuous frequency curve corresponding to an infinitely large hypothetical population.

It does not, of course, follow that a frequency-distribution when plotted as a frequency curve will yield a symmetrical curve, still less need it be that particular kind of symmetrical curve which is known as the *normal probability curve.* Nevertheless it has been found that the frequency-distributions not only for many physical traits like height but also for many mental traits like general intelligence do conform at least approximately to the normal probability curve. The mathematically minded will recognise that the equation of this curve may be written

$$y = y_0 \, e^{-\frac{x^2}{2\sigma^2}}$$

where y_0 is the value of the ordinate which passes through the origin O, e being the base of Napierian logarithms, and σ the standard deviation. Such a curve has two points of inflexion, marked I in Fig. 3. In other words, the curve changes its shape from convex to concave at these points, and the hori-

[1] R. A. Fisher, *Statistical Methods for Research Workers,* 1932, p. 37, 4th ed.

zontal distance of these two points from the summit is an excellent indication of the variability of the measures. This distance is denoted by σ and is known as the standard deviation. The method of calculating σ is given in Chapter 23. For such a curve the ordinate joining the origin O to the summit measures the mean, median or mode, for it is evident that these three measures must coincide when the distribution is normal. Such a curve on either side continually approaches the horizontal axis which is an asymptote. In other words, the curve will only meet the axis at infinity when $x = \pm\infty$. But in practice it is only necessary to consider that portion of the curve which lies between $x = \pm 3\sigma$.

It can be seen from the table that over 68 per cent of all the measures are included in that portion of the curve which lies between $x = \pm\sigma$, over 95 per cent between $x = \pm 2\sigma$, and 99·73 per cent between $x = \pm 3\sigma$.

THE NORMAL PROBABILITY CURVE

Areas corresponding to units of standard deviation measured from the mean along the base-line:

x.	Area.	x.	Area.
0·0	0000	1·6	4452
0·1	0398	1·7	4554
0·2	0793	1·8	4641
0·3	1179	1·9	4713
0·4	1554	2·0	4773
0·5	1915	2·1	4821
0·6	2257	2·2	4861
0·7	2580	2·3	4893
0·8	2881	2·4	4918
0·9	3159	2·5	4938
1·0	3413	2·6	4953
1·1	3643	2·7	4965
1·2	3849	2·8	4974
1·3	4032	2·9	4981
1·4	4192	3·0	4986·5
1·5	4332	3·1	4990·3

Thus, between the mean and a point $x = 1{\cdot}5\sigma$ there lies 43·32 per cent of the total area under the curve.

THE PROBABLE ERROR

It is now easy to calculate what is known as the *probable error*, although in reality it is not an error, and if it were it would not be probable! It refers to the value of the deviation x which corresponds to an area of 25 per cent of the total area. By interpolation the value of x is approximately $0{\cdot}6745\sigma$. Thus in the normal probability curve the probable error is

$$\text{P.E.} = 0{\cdot}6745\sigma$$

By definition, then, a deviation of the magnitude of the P.E. is as likely to be exceeded as not, for if this distance is marked on either side of the mean, the corresponding area will include half of the total frequency. It is an easy exercise to calculate from the above table that the probability of the occurrence of a deviation twice as large as the P.E. is 1 case in 5·6; and for one three times as large is 1 case in 23·5, approximately.

It is clear that cases will often occur where a normal distribution would not be expected, *e.g.*:

(1) The number of measures might be so few that an *irregular curve* would result to which no importance could be attached.

(2) The curve might be *bimodal* owing to the presence of two distinct classes in the group, thus the frequency-distribution of the stature of a mixed group of men and women, or that for the marks when boys from "good" schools and boys from "poor" schools sit the same competitive examination. A bimodal curve might also occur if two examiners of unequal severity shared the work of marking so that half the candidates were marked by one and half by the other.

(3) In still other cases a *skewed curve* may result.

(*a*) The skewness is positive when the longer tail of the distribution is in the direction of high scores (Fig. 4). This

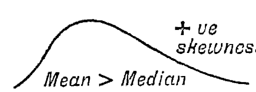

FIG. 4.

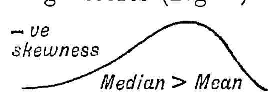

FIG. 5.

occurs when a test is too difficult for the candidates, or the marking has been too severe.

(*b*) The skewness is negative when the longer tail is in the direction of low scores (Fig. 5). This occurs when a test is too easy or the marking too lenient.

$Sk = \frac{Q_1 + Q_3 - 2Q_2}{Q}$ is Yule's measure of skewness, where Q_2 is the medium. (Cf. Chapter 23 for meanings of Q_1, Q_3 and Q.)

$Sk = \frac{3\ (\text{mean-median})}{\sigma}$ is Pearson's measure of skewness.

This is more exact and its significance, for the case of n measures, is tested by comparing it with its standard error (c.f. Chapter 23) which is $\sqrt{\frac{3}{2n}}$.

$Sk = P_{50} - \frac{1}{2}(P_{90} + P_{10})$ is Kelley's measure of skewness based on percentiles (see below). Its significance is tested by comparing it with its standard error which is $0{\cdot}59914 \frac{P_{90} - P_{10}}{\sqrt{n}}$.

PERCENTILES

It is often useful to know, not the actual measure of an individual, but his position with respect to his fellows. Thus to

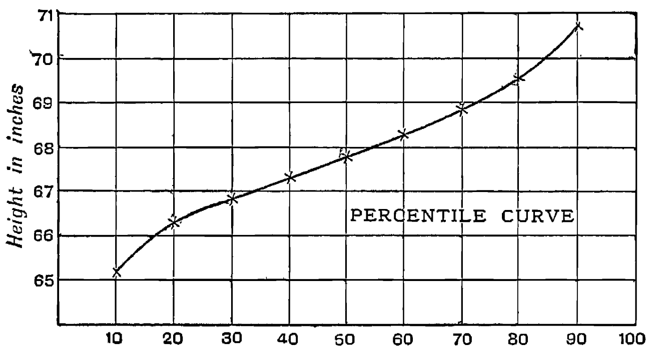

FIG. 6.

revert to the stature of Irishmen, what is the stature of a particular Irishman when just 50 per cent of all the Irishmen

measured are shorter than he is? The value of P_{50} in the table below supplies the answer. In other words, the percentile rank of this Irishman is 50 and his height is 67·75 inches, to the nearest quarter-inch. Cf. Fig. 6.

10 per cent of 345	= 34·5	$P_{10}=65+\frac{8\cdot5}{33}=65\cdot25$
20 ,, ,, ,,	= 69·0	$P_{20}=66+\frac{10}{58}=66\cdot25$
30 ,, ,, ,,	= 103·5	$P_{30}=66+\frac{44\cdot5}{58}=66\cdot75$
40 ,, ,, ,,	= 138·0	$P_{40}=67+\frac{21}{73}=67\cdot25$
50 ,, ,, ,,	= 172·5	$P_{50}=67+\frac{55\cdot5}{73}=67\cdot75$
60 ,, ,, ,,	= 207·0	$P_{60}=68+\frac{17}{62}=68\cdot25$
70 ,, ,, ,,	= 237·5	$P_{70}=68+\frac{47\cdot5}{62}=68\cdot75$
80 ,, ,, ,,	= 276·0	$P_{80}=69+\frac{24}{40}=69\cdot50$
90 ,, ,, ,,	= 310·5	$P_{90}=70+\frac{18\cdot5}{25}=70\cdot75$

THE RANGE OF INDIVIDUAL DIFFERENCES.—Attempts have been made to ascertain the magnitude of the differences which may exist in a given trait between the highest and the lowest score which appear in the sample tested. Bearing in mind the properties of the normal curve which have been mentioned in this chapter, it is easy to see that the ratio of the best measurement to the poorest[1] will depend on the number of individuals which have been tested, and also on the way these individuals have been selected. Thus the ratio would be greater for a large unselected sample of the general population than for a large sample of graduate students. This would hold not only for psychological traits such as Intelligence Quotients but also for educational abilities like speed of writing.

Further, as Wechsler[2] has pointed out, this ratio is at the

[1] A more stable ratio is that of the 75th percentile to the 25th, cf. Chapter 23.

[2] D. Wechsler, "On the Range of Human Capacities", *Proc. of 9th Intern. Congress of Psychol.* 1930, p. 476.

mercy of inequalities in the successive units of measurement as, for instance, when the successive problems in an arithmetical reasoning test are not of like difficulty. Starch[1] found this ratio in a class of only 36 pupils in Grade 8 to be 3·7 for speed of reading, and less than 2 for speed of writing. The present writer,[2] on testing twelve observers for light sensitivity, found that after a dark adaptation of 40 minutes, this ratio was 25; that is, one observer was 25 times as sensitive as another, and it seems reasonable to suppose that with a larger number of subjects larger differences would be found. If the magnitude of this ratio were calculated for industrial tasks, there are reasons for believing that high ratios would be comparatively rare; in fact, the average[3] would probably be under 2. Industry cannot afford a higher ratio. Unfortunately, in the inexorable process of reducing this ratio many employees fall by the wayside. Probably many of these falls could have been avoided if vocational guidance had been given and accepted. All honour then to those progressive firms which have instituted systems of psychological guidance so that an employee may be transferred from one department to another to the mutual gain of both himself and the firm.

ILLUSTRATIVE EXAMPLES

(1) What is the probability that an Irishman has a height between 72 and 73 inches?

By consulting the table it is seen that 10 Irishmen out of 345 are between 72 and 73 inches tall. Hence the probability is about 0·029, or 3 in 100.

(2) 200 candidates take an intelligence test. Assuming normal distribution, how many should receive the following grades: A (excellent), B (good), C (moderate), D (poor), E (very poor)? Assuming that the range of intelligence is the same in each grade, and remembering that the portion of the frequency curve which lies between $x = \pm 3\sigma$ includes 99·73 per cent of the

[1] D. Starch, *Educational Psychology*, 1920, p. 36.

[2] Ll. Wynn Jones, "A Method of Measuring Nyctopsis", *Brit. Journ. of Psychol.* April 1921, p. 316.

[3] C. L. Hull, *Aptitude Testing*, 1928, p. 35.

cases, it is only necessary to divide the length 6σ into 5 portions each of length $1{\cdot}2\sigma$.

From the table it is seen that 49·86 per cent of the cases fall between the mean and 3σ; and that 46·41 per cent fall between the mean and $1{\cdot}8\sigma$. Hence 3·45 per cent fall between 3σ and $1{\cdot}8\sigma$. Hence 7 candidates should receive the grade A. By similar reasoning it can be seen that there should be 48 B's, 90 C's, 48 D's and 7 E's.

QUESTIONS

1. Three problems were solved by 80 per cent, 70 per cent and 60 per cent of a large group. What is the relative difficulty of the problems, assuming them to be measures of a capacity which conforms to the normal curve?
2. What meanings may be given to the term "experience"? Give illustrations.
3. What meanings may be given to the term "intuition"?
4. What meanings have been given to the term "belief"? Give examples to illustrate each meaning.
5. "What conscience dictates to be done,
   Or warns me not to do,
   This, teach me more than hell to shun,
   That, more than heaven pursue."

   (POPE.)

   Do you consider that training in introspection tends to increase control over conduct?
6. Make a classification of mental differences. (Cf. refs. to C. Burt at the end of the chapter.)

REFERENCES

1. C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition*, 1923. (In the first eight chapters the three noetic principles are discussed in detail.)
2. C. Burt, "The Mental Differences between Individuals", *Brit. Assoc. Report*, 1923.
3. C. Burt, *The Measurement of Mental Capacities*, 1927.
4. J. McKeen Cattell, "Some Psychological Experiments", *Science*, January 1 and 8, 1926.

5. J. McKeen Cattell, *Psychology in America*, The Science Press, 1930.
6. J. C. Flugel, *A Hundred Years of Psychology*, 1933. (Gives a clear account of modern schools such as Configurationism or Gestalt, Behaviourism, "Hormic" Psychology, Psycho-analysis, and the "Factor" school.)
7. Psychologies of 1930, edited by C. Murchison, 1930. (An authoritative account by the leaders of the various schools.)

The Gestalt School

K. Koffka, "Introspection and the Method of Psychology", *Brit. Journ. of Psychol.* vol. 15, 1924.

K. Koffka, *The Growth of the Mind*, 1928.

W. Köhler, *Gestalt Psychology*, Amer. ed. 1929. Brit. ed. 1930.

Both Koffka and Köhler have criticised the method of introspection. The brilliant contributions by the leaders of the Gestalt school, including the Gestalt journal *Psychol. Forschung*, should be studied by the advanced student. After reading the following references the present writer is unable to range himself under the Gestalt banner, though as a pupil of Wundt and Spearman he cannot hope to be regarded as an unbiassed witness.

C. Spearman, "The New Psychology of 'Shape'", *Brit. Journ. of Psychol.*, vol. 15, 1925.

K. Koffka, "Psychologie der Wahrnehmung".
A. Michotte, "Sur la perception des formes".
E. Rubin, "Über Gestaltwahrnehmung".
F. Sander, "Über Gestaltqualitäten".
C. Spearman, "'Shapes' opposed to 'Wholes'".
8th Intern. Cong. of Psychol. Groningen, 1926.

H. Helson, "The Psychology of 'Gestalt'", *Amer. Journ. of Psychol.* vol. 36, 1925; vol. 37, 1926.

E. Rignano, "La Théorie de la forme de la nouvelle école psychologique allemande", *Scientia*, Sept., Oct., Nov. 1927.

E. Rignano, "The Psychological Theory of Form", *Psychol Rev.* vol. 35, 1928.

W. Köhler, "La Théorie de la forme" (Réponse à la critique de M. Rignano), and E. Rignano, "La Théorie de la forme" (Réponse à l'anti-critique de M. Köhler); both in *Scientia*, May 1928.

C. Spearman, "The Laws of Psychology", 9th Intern. Cong. of Psychol. 1929.

C. Spearman, "Formalism or Associationism?" *Brit. Journ. of Psychol.* vol. 19, 1929.

E. R. Jaensch, *Eidetic Imagery*, 1930.

O. A. Oeser, "Critical Notice of Köhler's Gestalt Theory", *Brit. Journ. of Psychol.* vol. 21, 1930.

B. Petermann, *The Gestalt Theory* (trans. by M. Fortes), 1932. (Contains a comprehensive bibliography of Gestalt literature.)

The Behaviouristic School

J. B. Watson, *Psychology from the Standpoint of a Behaviourist*, 1924. This book may be recommended for the more advanced student as an authoritative presentation of behaviourism. Although Watson criticises the method of introspection, his view is radically different from that of the Gestalt school.

The "Hormic" School

W. McDougall, *The Energies of Men: Fundamentals of Dynamic Psychology*, 1932.

CHAPTER 2

THE PERCEPTION OF SPEECH SOUNDS

MATERIALS.—A gramophone together with records of suitable speeches or recitations.

INSTRUCTIONS.—E informs the class that the gramophone needle will be placed very gently on the record chosen and lifted again after two or three words have been spoken, and that each S has then to write down all that he hears, and that this process will be repeated, using a different part of the record each time, until at least ten tests have been given. Before the experiment starts, preliminary practice should be given on another record, say, two or three trials. It is also advisable not to count the first trial on the record chosen, but to let it serve to make the Ss familiar with the quality of the speaker's voice.

Some check can be exercised on the words to be presented by fixing a graduated scale above the record in a plane parallel to it and passing through a point above its centre. The needle can then be placed opposite any desired point on the scale. It is important not to state what the title of the record is beforehand, and no particulars concerning it should be given. A rough estimate of S's performance in the test may be obtained by awarding a mark for each word correctly understood. A more exact method may be formulated if desired.

It should be distinctly understood that no claim is here made that this test can be employed as a group test in the sense that marks may be awarded to each S to indicate his efficiency, for without previous experimentation it cannot be ascertained how much the efficiency is dependent on S's exact position in the room. All that is now attempted is to use the test with a group with a view to discussing certain qualitative considerations.

DISCUSSION

The foregoing method of using the gramophone was explained to the present writer by Professor Spearman in 1914. The writer found at that time that ability in the test bears little relation to auditory acuity as usually determined by measuring the distance at which a constant sound can just be heard (cf. Chapter 13). The method actually employed, however, in order to measure the auditory acuity of the subjects was to present sounds of varying intensity by the use of a telephone receiver and a variable resistance. It must also be added that subjects whose hearing was abnormal were not included.

The interesting question thus arises: On what factors does success in the test depend? As already indicated, it is important not to state the title of the record or to give the name of the speaker. S would otherwise very probably gain higher marks in the test. The mere knowledge of the title may orientate S's mind in a certain direction and act as a powerful clue to the meaning of the phrases given. How much help may be obtained from a clue is shown very clearly by the following illustration taken from a somewhat different experiment. A selection called "La Chasse" was rendered on the gramophone to classes of graduate students in Leeds. They were then asked to give their impressions of the music. In one class no information concerning the music was given by E beforehand. Not a single S in this class made any reference to a hunt. But in another class the mere title—"La Chasse"—was given. As a consequence most Ss noted the hunting horn, the yelping of the hounds, the galloping of the horses, the losing of the scent, etc.

It is also obvious that the quality of the speaker's voice, his accent, style and peculiarities, will appeal to some more than to others, so that a subject will gain a comparatively higher score in one record than in another.

It should also be noted that the attitude of the subject in this test is somewhat different from that of the listener to an ordinary speech. The latter concentrates on the meaning while the task of the former is to concentrate on sounds shorn of

their context and, if possible, apprehend them as parts of sentences. In addition, S's knowledge of the contents of the speech or recitation, his previous acquaintance with the subject matter and his command over the vocabulary used, will exercise an influence over his score. But if we eliminate these factors by employing several records, an enormous difference in score between individuals will still be found, some scoring three times as many marks as others. It becomes therefore an interesting study to try and account for this. Some Ss appear to accept the suggestion given by a slight clue much more readily than others, so that many curious responses are made; while others are much more cautious and only attempt an answer when they are reasonably sure of being correct. It may be that mental inertia or perseveration plays a part in this test (cf. Chapter 18). If a teacher asks a question to which the answer is known to all, it will yet be found that some get the drift of the question sooner than others who may be equally intelligent. In any case, whatever factors are operative in the test described, it may well prove a good vocational test in the selection of telephone exchange operators.

Records which were found suitable for the test were Commander Shackleton's "Dash for the South Pole" and Mr. Asquith's "Speech on the Budget".

It is very instructive and sometimes amusing to note the errors made; thus "of social reform" appeared as "October record", "have told you before", etc.; "raise the price of food" appeared as "rays of life's book"; "ought to contribute to the na(tion)" appeared as "forced to continue to remain".

This experiment was placed early in the course, as it illustrates two very important truths which should be kept in mind from the outset. Firstly, there is no sharp-cut distinction between sensory perception and abstract thinking. In this experiment a series of auditory stimuli are presented to S and his task is to write down what they are. He has to listen and he has to think. In other words, although there is here a definite sensory basis, the end result is not a mere sensory perception but a more or less definite apprehension of a complex experience.

A trained S may perhaps venture to give some indication of how much more than bare perception is involved in this experiment. If so, he will be guided by the number and complexity of the relations which have to be cognised. As Spearman pertinently asks: "Who would rank as no more than sensory perception the understanding of Bach's masterpieces, where contrapuntal relations have to be cognised as themselves having inter-relations, and these again yet higher ones, in an unending pile upwards"? [1] Secondly, this experiment brings out the fact that the raw result of a psychological test may be conditioned by a plurality of causes, *e.g.* general mental ability, scholastic ability, auditory acuity, mental inertia, practice, age and position of S relative to the gramophone. If it is desired to study the effect of one of these factors, then it is obviously necessary to exercise some control over the others. Such questions require quantitative data and statistical analysis, and at a later stage of the course the treatment of simple cases will be illustrated by examples.

THE FORM OF PRESENTATION AND RECALL

An investigation by Exley [2] is somewhat germane to the topic discussed in this chapter. Materials were presented to a group of secondary school boys of average age twelve years six months, using three different methods: the visual method, where the material was presented pictorially at a cinema; the auditory method, when presented orally by an expert reader standing behind a screen; and the reading method, where the pupils studied hectographed sheets. With a view to study recall, the pupils wrote a composition on the morning following presentation and also twelve months later. Although the highest mean mark in the early recalls was obtained by the auditory method, with the visual method second and the reading method third, yet after a year the greatest permanence of recall was attached to the visual method, and least to the reading method.

[1] C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition*, p. 79.

[2] C. F. Exley, "A Survey of the Learning Process with special reference to Perception", M.Ed. Thesis, Univ. of Leeds, 1927.

the bottom, so that '2' is uppermost. I am going to show you the same pattern again for five seconds. After looking at it carefully, proceed as before on sheet '2', and introspect again. Then place '2' at the bottom so that '3' is at the top when the above procedure is repeated." This experiment should be done with simple patterns as shown in Fig. 7 and Fig. 8 and also with more complex patterns as in Fig. 9. It is not intended that the attempts with the simple patterns should be studied quantitatively, as some mature Ss succeed in reproducing the patterns after only one exposure. But an account should be given by each S of the method employed and the difficulties

R	B	Y	W
Y	R	W	B
B	W	R	Y
W	Y	B	R

FIG. 7.

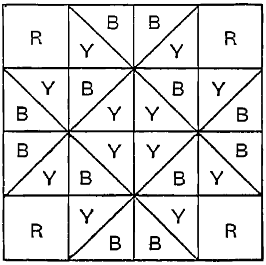

FIG. 8.

Y	B	B	B	Y	B	B	B	Y
B	W	R	R	W	R	R	W	B
B	R	W	B	B	B	W	R	B
B	R	B	R	R	R	B	R	B
Y	W	B	R	W	R	B	W	Y
B	R	B	R	R	R	B	R	B
B	R	W	B	B	B	W	R	B
B	W	R	R	W	R	R	W	B
Y	B	B	B	Y	B	B	B	Y

FIG. 9.

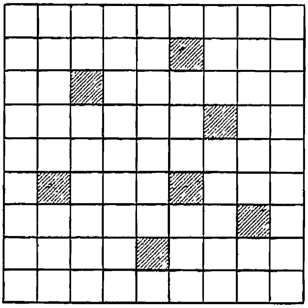

FIG. 10.

experienced. In order to get quantitative estimates with mature Ss, it is suggested that the more complex patterns be used, with appropriate blanks so that the number of exposures necessary for a correct reproduction may be noted.

DISCUSSION

A simple visual percept such as was obtained in these experiments, instead of furnishing a starting point from which to explain the laws of cognition, is something very complex, which needs such laws for its own explanation.[1] As in the last chapter it must again be emphasised that there is no profound gulf between sensory perception and abstract thinking; even in perception there is a continual finding of relations and the transition from perception to thought is a gradual one "by building upon the original sensory basis further (mostly relational) knowing of greater complexity".[2]

In a similar strain McDougall writes: "to know, to recognise, to be aware of, or conscious of, any object is to conceive it, even when our knowing is a perceptual knowing".[3] But even if the transition from perceiving to thinking is not well defined, the ancient distinction between "perceptual" and "intellectual" operations may nevertheless prove serviceable. Thus Strasheim, who has recently applied Spearman's principles in a new method of mental testing, found that the eduction of relations showed two stages. First, a perceptual stage where the relations are actually educed, but remain in intimate contact with the original fundaments, so that they will be effective in the future, only if the same or similar fundaments are supplied. Secondly, a stage where the relations have been definitely "abstracted", have become concepts so that they are effective for universal use, without the presence and help of perceived objects.[4]

SYMMETRY.—In the foregoing experiments it is at once apparent how an apprehension of the symmetry of the figure helps S to form an accurate percept so that he is soon able to draw a plan of the figure. This, of course, is done by apprehending the essential relations existing between the elements of the pattern.

[1] C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition*, p. 47.
[2] *Ibid.* p. 79.
[3] W. McDougall, *An Outline of Psychology*, p. 254.
[4] J. J. Strasheim, *A New Method of Mental Testing*, 1926, p. 141.

It is a noteworthy fact that young children even before they attain their fifth birthday are able to build patterns as complex as that shown in Fig. 9 when the appropriate cubes are available. One boy, aged four years nine months, interested himself for long periods in this way and would at once detect the lack of symmetry if someone turned over a particular cube.

When it is remembered how important a rôle is played, by various forms of what is rather loosely included under the term symmetry, in poetry, in music, in art, in mathematics and in the natural sciences, it becomes an important task for the educator to trace the development of symmetry. Without question symmetry has not received in schools the attention which it deserves. In a Report prepared by the Mathematical Association on *The Teaching of Geometry in Schools* some valuable references to symmetry will be found. Thus the Report recommends that "Symmetry should have a larger place in teaching than has been usual. It is a conspicuous quality of the things around us and of many geometrical figures with which we deal."

The apprehension of symmetry with its accompanying emotional excitement plays a large part even in the higher reaches of mathematical analysis. Thus Graham Wallas related how one mathematician, who had suddenly found a solution to a problem, could only account for it by thinking that an aesthetic feeling was excited.

It seems probable then that the study of symmetry will throw some light on the difficult question of special ability in mathematics.

In his study of "Psychological Tests of Mathematical Ability" Fouracre[1] found the above Symmetry test to give promising results, as it correlated positively with all the other twelve tests given, both for Elementary and Secondary schoolboys. Especially highly did it correlate with his geometrical tests. Using partial correlation coefficients (cf. Chapter 23) the Symmetry test correlated with the pool of the geometrical tests:

0·71 (probable error 0·107) for Secondary schoolboys
0·80 (,, ,, 0·11) ,, Elementary ,,

[1] L. Fouracre, M.Ed. Thesis, Univ of Leeds, 1925. Cf. also *The Forum of Education*, vol. 4, No. 3, November 1926.

The geometrical tests employed were the Thurstone Hand test, Thurstone Spatial Relations A, Thurstone Spatial Relations B, obtainable from the C. H. Stoelting Co., Chicago, and a Geometry test devised by Fouracre himself.

NON-SYMMETRICAL PATTERNS.—In order to show how difficult the task becomes when there is a marked lack of symmetry in the pattern, the colour cube box containing nine cubes to the side was arranged so that the surface is all white except the seven red squares (represented as black in Fig. 10). This form of the test bears a similarity to the Spot Pattern test of McDougall.[1] The great advantage of the test is that it can be varied at will and made of any required degree of difficulty. It suffers from one drawback, however, since an ingenious student may be able to evolve a method by means of which it is possible to reproduce the figure after only one exposure, unless this is short, say, under five seconds. What is the method?

An alternative task[2] for the study of visual perception is that of presenting an irregular line consisting of some eight segments alternately straight and curved, each exposure lasting ten seconds. This task bears some relation to the activity of a child learning to write. A portion of a map or a piece of mechanism or even a letter of the alphabet may not be correctly reproduced by a pupil because his perception of them has never been exact. His teacher may think it a case of weak memory whereas it may be more accurately regarded as a case of failure to educe those relations which are the basis of exact perception.

The particular forms of the above experiments were chosen for various reasons:

(1) A test can be completed in a short time. This would not be the case for such a complex task as learning to write.

(2) The method is genetic and hence offers the great advantage to the student of noting any change in conscious experience as a result of practice.

[1] Cf. G. M. Whipple, *Manual of Mental and Physical Tests*, vol. 1, p. 290.

[2] Cf. C. H. Judd and D. J. Cowling, "Perceptual Learning". *Psychol Rev.* Monograph Supplement, vol. 8; F. N. Freeman, *Experimental Education*, 1916. p. 25.

(3) Drawing ability is practically eliminated, for no S would lack the skill to draw the figure if it was correctly perceived. Thus drawing a figure in which the elements present no technical difficulty is an excellent means of testing the accuracy of the percept. Even in drawing, the perceptual element is now recognised to be more important than the motor element. Also in handwriting the perception of form precedes reproduction and is made more precise by the effort at reproduction.[1]

(4) The task has a definite perceptual basis.

(5) Finally, although the tests lend themselves to quantitative measurement, they serve here as exercises in introspection and illustrate the different methods and attitudes of the Ss towards the tasks. Some tend to concentrate on the figure as a whole while others seem to concentrate more on a part at a time. Some vary their methods more than others during learning. Sometimes artificial associations or mnemonics are employed in order to master the task. Some, at first, adopt a somewhat passive attitude in looking at the figure, *e.g.* one S observed: "At the first exposure there was little active exploration of the figure but a mere passive reception of a general impression. On proceeding to draw the figure, the uselessness of this passive attitude became apparent."

Most, however, are actively employed in exploring the figure, in selecting relevant items and in controlling the whole process of learning. Thus Woodworth[2] calls learning a reactive rather than a receptive process, and F. C. Bartlett[3] makes "the effort after meaning" the mark of perceiving, and points out that an apparently arbitrary mixture of lines and curves most effectively resists the "effort after meaning", while anything symmetrical is very quickly noticed. It was also found that simple and familiar material tended to be perceived as a whole, but unfamiliar, disconnected or relatively detailed material was, as might be expected, perceived more analytically.[4] He who runs may read the pedagogical corollaries! There are also

[1] F. N. Freeman, *Experimental Education*, p. 25.
[2] R. S. Woodworth, *Dynamic Psychology*, p. 107.
[3] F. C. Bartlett, "An Experimental Study of Some Problems of Perceiving and Imaging", *Brit. Journ. of Psychol.* vol. 8, No. 2, 1916, p. 232.
[4] *Ibid.* p. 249.

differences with respect to imagery exhibited in the foregoing tests: some rely more than others on visual imagery and may be able to see the figure "in their mind's eye" when it is no longer on view; others may rely more on kinaesthetic[1] imagery, and their learning may be assisted by movements of arms or fingers, lips or glottis. Some may use both visual and kinaesthetic and some may use neither. Auditory imagery and verbal imagery may also be of help. By this is not meant that individuals may be divided into visiles, audiles and motiles as was presumed by some to be the case. It is more correct to say that most people use all forms of imagery, but that certain kinds of imagery play a more prominent part than others in the case of some individuals. The subject of imagery is, however, one of the most difficult which psychologists have to contend with, and it is not therefore surprising to find no unanimity as to the nature and function of imagery. Such a topic then should appear later on in the course, if it should appear at all.

MECHANICAL AIDS IN EDUCATION

Owing to the advent of the gramophone, wireless and the cinema to the schools, new problems for teachers and psychologists have arisen. Visual and auditory stimuli can be presented to pupils to a degree hitherto unattainable. The educational value of each mechanical aid, in relation to each school subject, must be studied in detail. When, and to what extent, should gramophone records be employed in the study of modern languages? How helpful is the film in the study of history or geography?

A most comprehensive investigation has recently been completed by Dr. Frances Consitt[2] on the value of films in history teaching. The results showed that the film gives life to the past by portraying life in movement, and by giving a background with full detail; it arouses interest that stimulates the children to further mental effort; it stimulates imagination; it helps them to assimilate so that they gain both in grip and atmosphere; it

[1] C. S. Myers, *Textbook of Experimental Psychology*, Part 1, chap. 12.
[2] F. Consitt, "The Value of Films in History Teaching", Ph.D. Thesis, Univ. of Leeds, 1930. Cf. reference at end of chapter.

helps them to remember; it forces them to find their own words to express opinions and describe scenes, not merely to borrow those of the teacher or textbook; it presents a point of view to the children in addition to that of the teacher and the textbook; and finally it affords pleasure.

References to some of the most important investigations on films appear at the end of the chapter.

SPEARMAN'S QUANTITATIVE PRINCIPLES

The five fundamental quantitative laws of cognition formulated by Spearman may be illustrated by experiments such as those described in this chapter.

LAW OF ENERGY.—The first quantitative principle is the Law of Limited Mental Energy or the Law of Constant Output:

Every mind tends to keep its total simultaneous cognitive output constant in quantity, however varying in quality.[1] In other words, the "span of consciousness" is limited in extent. A simple illustration is to expose a number of dots for a tenth of a second by means of a short exposure apparatus known as the tachistoscope. If the number does not exceed seven, many adults are able to apprehend them as seven, but let there be eight or more and few would be able to vouch for the correct number, unless they can be grouped into higher units. If, however, they are arranged in groups of seven or less, it is still possible to apprehend up to seven of such groups. Thus as many as forty-nine dots may be apprehended at one glance. The bearing of this principle in experiments such as those described in this chapter on the apprehension of coloured squares is obvious.

LAW OF RETENTIVITY.—The second quantitative principle is the Law of Retentivity:

The occurrence of any cognitive event produces a tendency for it to occur afterwards.[2]

Thus in the experiments with coloured patterns, reproduction is facilitated as the number of exposures is increased.

[1] C. Spearman, *The Nature of Intelligence*, 1923, p. 131.
[2] *Ibid.* p. 132.

Spearman points out that this tendency to reoccur may manifest itself in two different ways so that we should more properly talk of two laws of retentivity instead of one. The first is the—

Law of Dispositions: Cognitive events by occurring establish dispositions which facilitate their recurrence.[1]

A special case of the law of dispositions is the *Law of Association*: Cognitive events by accompanying each other establish dispositions to do so thereafter.

The second law of retentivity is the—

Law of Inertia: Cognitive processes always both begin and cease more gradually than their (apparent) causes[2] (cf. Chapter 18).

The third quantitative principle is the—

LAW OF FATIGUE.—The occurrence of any cognitive event produces a tendency opposed to its occurring afterwards.[3]

At first glance this principle appears contradictory to the principle of retentivity and its exposition is reserved for Chapter 19.

The fourth quantitative principle is the—

LAW OF CONATIVE CONTROL.—The intensity of cognition can be controlled by conation.[4]

Thus the subject's success in reproducing the coloured squares undoubtedly depends on how well he "concentrates" on or "attends" to the task.

Finally there is the fifth quantitative principle or the—

LAW OF PRIMORDIAL POTENCIES.—Every manifestation of these four principles is superimposed upon certain primordial but variable individual potencies such as those of age, sex, heredity and health.[5]

Thus although the span of consciousness is constant for a

[1] C. Spearman, *The Nature of Intelligence*, 1923, p. 133. Cf. also C. Spearman, *The Abilities of Man*, chap. 16, 1927.

[2] *Ibid.* p. 133; *The Abilities of Man*, chap. 17.

[3] *Ibid.* p. 134; *The Abilities of Man*, chap 18.

[4] C. Spearman, *The Nature of Intelligence*, p. 135; *The Abilities of Man*, chap. 20. Cf. E. H. Wild, "Influences of Conation on Cognition", *Brit. Journ. of Psychol.* vol. 18, Nos. 2 and 3, 1927–28.

[5] C. Spearman, *The Nature of Intelligence*, p. 136; *The Abilities of Man*, chaps. 21 and 22.

given individual at a given time, yet on account of such processes as growth and decay the span for a youth of eighteen is greater than that of a child of eight or of the veteran of eighty. The researches of Thorndike and those of Miles have shown, however, that the deterioration due to age is much less than popularly supposed.[1]

Here, then, are eight principles—three qualitative and five quantitative—which must suffice for the study of cognition in general, and which particularly illumine processes of learning such as have been discussed in this chapter.

QUESTIONS

1. By what method did you reproduce Fig. 1 ? Can you think of a more effective method?
2. Were you aided at all in these experiments by images? If so, in what way?
3. From your experience in this experiment frame instructions which are likely to increase the efficiency of S in reproducing designs and drawings in general and not merely those of a specific nature.
4. "Differences of quality, of colour, are of relatively small practical importance in visual perception; pattern predominates vastly" (W. McDougall, *Outlines of Psychology*, 1924, p. 233). Discuss this with reference to the use of black-and-white drawings, silhouettes and coloured pictures with children of various ages.

REFERENCES

1. C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition*, chap. 15.
2. F. N. Freeman, *Experimental Education*, 1916, chap. 2.
3. F. C. Bartlett, "An Experimental Study of Some Problems of

[1] Cf. E. L. Thorndike, E. O. Bregman, J. W. Tilton and E. Woodyard, *Adult Learning*, 1928; W. R. Miles, "Measures of Certain Human Abilities throughout the Life Span", *Proc. Nat. Acad. Sci.* vol. 17, 1931; C. C. Miles and W. R. Miles, "The Correlation of Intelligence Scores and Chronological Age from Early to Late Maturity", *Amer. Journ. of Psychol.* vol. 44, 1932; W. R. Miles, "Abilities of Older Men", *Personnel Journ.* vol. 11, 1933; W. R. Miles, "Age and Human Ability", *Psychol. Rev.* vol. 40, 1933.

Perceiving and Imaging", *Brit. Journ. of Psychol.* vol. 8, Part 2, May 1916.

The following references to investigations with Films may be recommended:

1. *Visual Education*, ed. by F. N. Freeman, 1924.
2. G. Révész and J. F. Hazewinkel, "The Didactic Value of Lantern Slides and Films", *Brit. Journ. of Psychol.* vol. 15, 1924.
3. *The Cinema in Education*, ed. by Sir James Marchant, 1925, gives the results of S. J. F. Philpott's experiments at Univ. Coll., London.
4. *The Educational Screen*, November 1928; October 1929.
5. D. C. Knowlton and J. W. Tilton, *Motion Pictures in History Teaching*, 1930, contains the results of experiments at New Haven.
6. F. Consitt, *The Value of Films in History Teaching*, 1931, *being the Report of an Enquiry conducted under the Auspices of the Historical Association with the Aid of the Carnegie United Kingdom Trustees.*

CHAPTER 4

THE TACTUAL PERCEPTION OF FORMS

MATERIALS.—Cornell Form-Board.[1] Stop-watch. Large cloth or cardboard to cover the Form-Board and blocks. Blindfold for S.

METHOD.—The apparatus should not be seen by S until the testing is completed. Let S stand before a table of convenient height with the form-board squarely in front of him. E should learn by heart the scheme of numbering the blocks and their corresponding holes. The numbering runs from 1 to 10, beginning at the upper left-hand corner and proceeding from left to right. E should then have the blocks arranged on the table to the right of S if he is right-handed, so that they may be presented to S in the order 5, 8, 3, 1, 7, 9, 6, 2, 4, 10. When S has been blindfolded and led to the position in front of the board, instruct him thus: "On the table before you there is a board in which are a series of ten holes of different shapes. I shall hand you, one at a time, a series of blocks of wood that just fit these ten holes. Put each block in place as quickly as you can. A new block will be placed for you just here (indicating a point on the table just at the right of the board) as soon as you have placed the one before it. You must place each block properly before you pick up the next one. Do you understand? Ready! Go!"[2]

At this signal E starts the watch and records the time taken to place all the blocks correctly. S is allowed to handle the block and examine the board with his fingers in any way that he wishes. Note the exact method used and record any points of interest in S's way of accomplishing the task.

A qualitative record which shows what precisely S does with each block should be made in the following manner: as S attempts to place a given block, E records against the number

[1] Obtainable from the C. H. Stoelting Co., 424 North Homan Ave., Chicago.
[2] G. M. Whipple, *Manual of Mental and Physical Tests*, vol. 1, p. 301.

of that block the numbers of the holes into which S tries to place it. If a block is correctly placed without any false moves, no number will appear; otherwise, the series of numbers recorded against a given number indicates the full series of trials until the block has been correctly placed, *e.g.*

5 – 9 – 3.

Here S first tried to place block No. 5 in hole No. 9, then in hole No. 3, then correctly in hole No. 5. When the last block is correctly placed, cover the board carefully with the cloth, remove S's blindfold and let him try to sketch the positions of the forms and their shapes. Restore the blindfold and repeat the test precisely as before. Make a third similar trial or continue the test in this manner until S is able correctly to reproduce all the forms in their proper positions. Note the number of trials made and plot a curve of time taken in relation to number of trial.

In demonstrating this test with the Cornell board to a large class, it is advisable to choose an S who does not mind being the cause of mirth, for as a rule, every member of the class is convinced in his own mind that he himself could do better. The shapes used in the Cornell board are similar to those of the Goddard board, but as the shapes in the Cornell board fit into a series of removable blocks, their relative position in the tray may be changed at will.[1]

Should E desire it, a new arrangement of the holes inside the tray may be presented to S at each attempt, but usually the same arrangement of shapes is used throughout, so as to test S's ability to learn. It is obvious that in the latter case the reduction in time at each successive attempt will be more pronounced.

DISCUSSION

> (The children) come to discriminate between small forms varying very slightly such as corn, wheat, and rice. They are very proud of seeing without eyes, holding out their hands and crying, "Here are my eyes!" "I can see with my hands!"
>
> M. Montessori, *The Montessori Method*, p. 190

In experiments with the form-board care must be taken in interpreting the results. Before drawing any conclusions from

[1] This improvement was suggested to Professor Whipple by Mr. D. Kennedy-Fraser.

the nature of the errors made by S, it is essential not only to note what errors are made, but also to secure S's introspection, *e.g.* S may rely on his memory of where a particular block is situated but actually attempt to place it in an adjoining position as, owing to his being blindfolded, his sense of location is not very exact. Here, as always, so much depends on the directions given. If S is merely told to place the blocks in position as quickly as possible, he may think it profitable to attempt to put a block in a particular hole after only a partial exploration. If, however, he is told that errors, *i.e.* attempts to place a block in incorrect positions, are counted against him, then his attitude is appreciably altered.

Data secured by Whipple from a small number of advanced university students give for three successive trials without sight of the board averages of 140, 120 and 69 seconds respectively. "When the attempt was made to sketch the board after each trial, it was found that no S could accomplish this after the first trial, whereas about one-half located the blocks correctly and sketched their forms with but trivial errors after the third trial. Decided individual differences are discovered in capacity to identify forms tactually. It is possible for S to learn to place the blocks rapidly by touch and yet have extremely erroneous ideas of their actual form."[1]

THE FORM-BOARD: VISUAL METHOD

In the tactual method already described, perception of form is a prime factor even with older children and adults, but in their case if they are allowed to see the board, performance in the test is then largely conditioned by speed and co-ordination of movement. But for many children under ten years of age and for feeble-minded children the perceiving of the forms is not facile even when they are not blindfolded. For such cases the blocks are placed in a standard order at the right side of the board and a slight modification of the wording of the instructions already given will enable E to employ the visual method. Norms for children of ages from five to fourteen years

[1] G. M. Whipple, *loc. cit.* p. 303.

are given by Miss F. Gaw in Report No. 31 of the Industrial Fatigue Research Board.

HISTORICAL DEVELOPMENT OF THE FORM-BOARD

At the beginning of the last century Itard[1] employed cardboard forms of various shapes and colours when attempting to educate the "Wild Boy of Aveyron". Fifty years later another French physician, Séguin,[2] published an account of methods for training feeble-minded children which included the systematic use of form-boards. Early in the present century Montessori successfully used Séguin's material in teaching normal children. About this time also, form-boards began to be used, not only for teaching purposes but also for testing. Thus Norsworthy[3] in 1906, using the form-board with a group of mentally defective children, tested their form perception and rate of movement and their learning capacity.

PERFORMANCE TESTS

During the last twenty years many other tests of a non-verbal type besides the form-board test have been employed for testing the "intelligence". Such tests are now known as Performance Tests, and "intelligence" is measured by the performance of some act and not by an oral or written response. But it follows as a corollary of the Two-Factor Theory[4] that no single performance test would furnish an accurate measure of general mental ability. A series of different performance tests would be required. Just as different linguistic tests are employed in all the well-known scales of intelligence, whether individual or group. Accordingly Pintner and Patterson devised a comprehensive Performance Scale. The uses of such a scale may be summarised as follows:

(1) To test foreigners who speak little if any of the language in which the test is given. Thus in America they have been

[1] J. Itard, *De l'éducation d'un homme sauvage*, 1801 and 1807.

[2] E. Séguin, (1) Traitement moral, hygiene et éducation des idiots, 1846; (2) Idiocy and its Treatment by the Physiological Method, 1866.

[3] N. Norsworthy, "The Psychology of Mentally Deficient Children", *Archives of Psychology*, No. 1, 1906.

[4] Cf. Chapter 8.

extensively used with the foreign-born and with non-English-speaking subjects. They are also of use with illiterate persons.

(2) To test subjects with physical abnormalities which handicap them in a linguistic test, *e.g.* the blind, the deaf. Pintner and Patterson's Performance Scale was specially designed to test the deaf. By its use the deaf child was found to be retarded in comparison with the hearing child, although not to the same extent as was indicated when a linguistic scale was employed. Drever and Collins, however, who have recently tested deaf and hearing children by means of performance tests, do not find a significant retardation. But they add, "Whether retardation will appear as our investigation proceeds and our numbers accumulate remains to be seen".[1] In comparing the results obtained by different investigators with different groups of deaf subjects, it is important to get comprehensive data, including the medical history, for each subject in order to ascertain if the different groups are directly comparable.

(3) To test various classes of subjects who exhibit various degrees of abnormalities, or at least idiosyncrasies, which are more complex in character than blindness or deafness, *e.g.* the "verbalist" who is said to be an individual whose capacity to use language is much above his ability in other respects, and it is well to ascertain if any such can be found.

Similarly with the tongue-tied of whom the converse would be true. Other cases are children presenting special scholastic or disciplinary problems; many different kinds of cases of aphasia where symbolic thinking and expression is affected; mental defectives and cases of border-line intelligence; delinquents and psychopaths of various types.

(4) To test normal subjects. Thus they may be used to supplement the result of a linguistic test such as the Binet Scale, *e.g.* a vocational counsellor is often materially helped in his prognosis by their use. They may also be used with young children as a substitute for the Binet Scale on account of their greater interest and independence of linguistic response.

At the same time the following passage by Sir Henry Head

[1] J. Drever and M. Collins, *Performance Tests of Intelligence*, 1928, p. 18.

might suggest that linguistic tests are superior in the long run to performance tests for assessing general mental ability: "The power to make use of words was acquired at a late stage in the evolutionary history of man. Before he learnt to formulate his thoughts in speech, he was capable of a wide range of accurate discrimination, which could express itself in action only. Progressive development of the higher cerebral centres led to increased discriminative powers; and speech, one of the highest and most recently acquired functions, was from the first concerned with expressing relations."[1] But the recent results obtained with non-verbal tests, described in Chapter 8, must be first considered.

Caution must also be exercised in view of the recent results obtained by the National Institute of Industrial Psychology[2] which seem to indicate that in the case of the older children success in a performance test may depend largely on a specific or group factor which has not yet been sufficiently developed in the case of the younger children. Thus the former practice of calculating a mental age based on performance may, in the case of older children, be open to criticism.

QUESTIONS

1. Contrast the tactual method of using the form-board with the visual method when used—
   (*a*) with children under seven years of age;
   (*b*) with adult subjects.
2. A girl, aged three, blindfolded, is given the bricks and cubes of Froebel and is told to place the cubes to the right and the bricks to the left. She succeeded in recognising the objects before feeling their contours, according to Montessori (cf. *The Montessori Method*, p. 189). Discuss this.
3. Give an introspective account of how you yourself recognise the forms which are used in the form-board test.
4. Name various classes of persons for whose examination Performance Tests are necessary.

[1] H. Head, "Disorders of Symbolic Thinking and Expression", *Brit. Journ. of Psychol.* vol. 11, Part 2, p. 179.

[2] Report No. 53 of the Industrial Fatigue Research Board.

5. What is the relation of spatial cognition to "g" (Spearman's General Factor)? How would you proceed to determine whether *all* spatial cognition possesses a common group factor? (C. Spearman, *The Abilities of Man*, pp. 176, 177, 228). This question is intended for the more advanced student.

References

1. H. H. Goddard, "The Form-board as a Measure of Intellectual Development in Children", *The Training School Bulletin*, vol. 9, 1912.
2. W. F. Dearborn, J. E. Anderson and A. C. Christiansen, "Form Board and Construction Tests of Mental Ability", *Journ. of Educ. Psychol.* vol. 7, 1916.
3. R. Pintner and D. Patterson, *A Scale of Performance Tests*, 1917.
4. G. O. Ferguson, "A Series of Form Boards", *Journ. of Exp. Psychol.* vol. 3, 1920.
5. W. F. Dearborn, E. A. Shaw and E. A. Lincoln, "A Series of Form Board and Performance Tests of Intelligence", *Harvard Monog. in Educ.* Series 1, No. 4, 1923.
6. Frances Gaw, *Performance Tests of Intelligence*, Industrial Fatigue Research Board, Report No. 31, 1925.
7. H. Binns, "The Discrimination of Wool Fabrics by the Sense of Touch", *Brit. Journ. of Psychol.* vol. 16, 1926.
8. F. L. Wells, *Mental Tests in Clinical Practice*, 1927.
9. A. F. Bronner, W. Healy, G. M. Lowe and M. E. Shimberg, *A Manual of Individual Mental Tests and Testing*, 1927.
10. J. Drever and May Collins, *Performance Tests of Intelligence*, 1928.
11. F. M. Earle, M. Milner and others, *The Use of Performance Tests of Intelligence in Vocational Guidance*, Industrial Fatigue Research Board, Report No. 53, 1929.

CHAPTER 5

MEMORY

A. RECALL AND RECOGNITION

MATERIALS.—E decides on the kind of material to be used, whether simple nouns such as field, cloud, etc.; nonsense syllables such as tiv, gop, etc.; geometrical forms; pictures of common objects such as knife, horse, etc.; or proverbs. Proverbs are more suitable for adults than for children. A list of one hundred suitable proverbs is given by E. M. Achilles in "Experimental Studies in Recall and Recognition", *Archives of Psychol.* No. 44, 1920.

More than one kind of material should be used. The form of the instructions will differ according as the test is a group or an individual one. The following arrangement is suggested for a group test with simple nouns.

E should prepare a list of 75 simple nouns such as cloud, field, etc., 25 of these should be printed in single column on Sheet A, another 25 in single column on Sheet B while on Sheet C should appear in two columns the remaining 25 mixed with the 25 which appeared on Sheet B.

First Experiment: Recall

INSTRUCTIONS.—Each S is given a copy of Sheet A and is allowed 50 seconds to study the sheet with a view to reproducing on a piece of paper as many words as possible. At the end of 50 seconds the sheet must be placed out of sight, and 30 seconds later, S writes down in any order as many of the 25 as he can remember.

Note.—In order to compare Ss as to their powers of complete recall, an individual test with this material could be arranged. After each presentation as above, S writes down as many as he can remember of the 25 words. This is repeated until S succeeds

twice in succession in reproducing all the 25. The number of presentations is noted.

Second Experiment: Recognition

INSTRUCTIONS.—Each S is given a copy of Sheet B and is allowed 50 seconds to study the sheet. It is then placed out of sight, and 30 seconds later, each S looks at Sheet C and marks with a cross all the words which he recognises as having appeared on Sheet B. The score is the number right minus the number wrong.

B. REPRODUCTION OF UNIQUE EXPERIENCES AND REPRODUCTION AFTER REPETITION

Reproduction of Unique Experiences: First Experiment

MATERIALS.—The Story of "The Two Shepherds" as given in the Manual of Directions for the Otis Group Intelligence Scale, together with a copy of the Otis Group Intelligence Scale. Advanced Examination, Form B, published by the World Book Company, Yonkers-on-Hudson, New York.

Directions.—The Story of "The Two Shepherds" should be read by E to the class according to the instructions given in the Otis Manual. Then the thirty questions on the story which appear in the Advanced Examination, Form B, should be answered by each S according to the instructions given.

Second Experiment

MATERIALS.—A copy of Ballard's Silent Reading Test, as given in Ballard's *Mental Tests*, p. 147, should be available for every S, together with the accompanying Completion Test.

DIRECTIONS.—Ss are allowed three minutes to read the story and are informed that they will afterwards be tested to see how much they have remembered. At the end of the three minutes the copies of the story are collected and copies of the completion test are distributed, and S's attempts to supply the missing words are marked according to the directions given in Ballard's *Mental Tests*, p. 149.

Third Experiment

MATERIALS.—Coloured pictures: "A Disputed Case" and "Australians". These pictures may be obtained from the C. H. Stoelting Co., 424 North Homan Avenue, Chicago, and the corresponding Interrogatories may be found in Whipple's *Manual of Mental and Physical Tests*, vol. 2, pp. 27-29.

INSTRUCTIONS.—"I am going to show you a picture in order to test your memory. Afterwards I shall want you to tell me what you have seen and I shall ask you questions about many little details." The picture is exposed for 20 seconds to the group but E does not specify the time of exposure. At the end of the twenty seconds the picture is removed and the S's are directed as follows: "Write down everything you saw; describe it so clearly that if I had never seen the picture I should know all about what was on it". Proceed next with the interrogatory. E asks the questions in the order given and S writes his replies.

When this test is used as an individual test with children, both the narrative and the replies to the interrogatory may be given orally by the child and recorded verbatim by E.

The results obtained with one picture may be compared with those obtained with the other.

Reproduction after Repetition: First Experiment (Symbol-Digit Test)

MATERIALS.—A copy of the Symbol-Digit Test as given in the National Intelligence Tests, Scale A, Form 1, together with its accompanying Manual of Directions as supplied by the World Book Company, Yonkers-on-Hudson, New York.

DIRECTIONS.—The test is carried out according to the directions given. At the beginning Ss have to refer to the key at the top, but the quicker they master the necessary associations the sooner they will be able to dispense with the key and the sooner they are able to finish the task.

Second Experiment (Modified Code Test)[1]

MATERIALS.—The words on the sheet described below should be copied by each S according to the directions given.

[1] I have thus modified a famous Code Test so as to make it more suitable for use as a group test with children.

DIRECTIONS.—On one side of a sheet of paper the following words should be written by each S:

ache	fife	badge
deaf	hide	cafe
beige	ice	dab

On the other side the following words in eight rows:

acid	he	fib	cage
chef	big	dag	fid
gib	face	bed	hag
head	gibe	cab	if
bah	chief	dice	gad
fed	bag	chaff	hie
bice	ah	high	fade
fag	bid	beg	chid

Ss then turn the paper over so that they are looking at the side containing only three rows.

E puts on the blackboard the following diagram:

a	d	g
b	e	h
c	f	i

FIG. 11.

and explains the code, namely, that "a" is _|, "b" is ‾|, etc., and that S guided by the diagram should write "ache fife badge" in code under the third line. The diagram is then rubbed off and Ss write "deaf hide cafe" in code underneath the last line, and finally "beige ice dab" in code in the same way. Ss are warned not to draw the diagram. When all have finished E gives the signals "Ready! Go" and Ss turn the sheet over and write as much as they can of the other eight lines. This side of the sheet only is marked and the time allowed is three minutes.

Third Experiment (Vocabulary Test)

Twenty cards are prepared, on each of which is printed an English noun and its Hindustani equivalent, and of a size big

enough to ensure their being seen by the entire class. They are exposed consecutively at two seconds' intervals. After the whole list has been exposed twice, the Hindustani words are shown in a prearranged order and Ss attempt to write down their English equivalents, and the attempts are collected. The same performance is repeated until one member of the group has almost succeeded in reproducing all. Sur[1] employed the following list:

BASKET	TOKRI	LIGHT	ROSHNI
HOUSE	MEKAN	DANCE	NACH
TAILOR	DURJI	COLOUR	RUNG
CLOUD	BADUL	BOAT	KISHTI
GLASS	SHISHA	HONEY	SHEHUD
WATER	PANI	CHAIR	KOORSI
CHURCH	GIRJA	STICK	DUNDA
LETTER	KHUT	BOOK	KITAB
BUSH	JHARI	MOUSE	CHOOHA
GOAT	BUKRA	RIVER	NUDI

More difficult is the following English-Arabic list:

SHEPHERD	GHANNAM	WEATHER	DINYA
CLEVER	SHATIR	SYSTEM	TARTIB
TWENTY	ISHRIN	ENGINE	WABUR
PREFER	FADDAL	WASHING	GHASIL
ASLEEP	'NAYIM	CONTENT	MABSUT
ERRAND	MISHWAR	THREATEN	HADDID
COUNTRY	BILAD	CONSCIENCE	ZIMMA
MISTAKE	GHALTA	BANDAGE	RUBAT
COMPLAINT	SHAKWA	EMPLOY	KHADDIM
EXPLAIN	FASSAR	CONQUER	GHALAB

DISCUSSION

In 1879 Ebbinghaus commenced what was probably the first experimental study of memory. He it was who devised the methods of complete mastery and of savings (cf. end of this

[1] R. K. Sur, "An Experimental Investigation of Memory of School Children with special reference to Bergson's Theory", M.Ed. Thesis, Univ. of Leeds, 1929.

chapter). He also invented the nonsense syllable technique which enabled him to get learning lists which were homogeneous and which avoided associations already formed as far as possible. Thus zab, lem, gir, might be suitable examples.

Many hundreds of investigators have since followed in his wake, probably a greater number than in any other single topic with the exception of that of "intelligence". Why such activity? There are probably two reasons. In the first place, memory appeared to have such vital connexion with the conditions of learning.

In the second place, it appeared fairly easy in the field of memory to exercise that experimental control so dear to the heart of the "brass instrument psychologist". What is easier than to expose a definite number of nonsense syllables for a definite time? There are, however, a number of possible pitfalls, so that rigorous control must be exercised before valid conclusions can be drawn. It must be emphasised that:

(1) Different individuals vary widely in their methods of memorising and unless it is ascertained what these may be, it is not safe to draw general conclusions.

(2) The stage of practice of the subjects must be ascertained and allowed for, and it is essential that Ss should first undergo preliminary practice.

(3) McDougall[1] has pointed out that many investigators in the field of memory made their Ss form associations by frequent repetitions but then unfortunately interpreted their results as holding true of memory in general.

(4) There is the peculiar danger of neglecting the factor of "g"[2] and the complementary danger of regarding memory as a faculty in spite of the fact that the doctrine of faculties is supposed now to be discarded. For ages memory held sway as one of the "big four", perception, intellect and imagination being the other three. "Such faculties, although now having to be disposed by us from their unbroken reign in psychology, nevertheless still do possess and will retain no little subsidiary value. They provide at least a preliminary sifting of a large

[1] W. McDougall, *Body and Mind*, chap. 24.

[2] Cf. Chapter 8.

quantity of facts. And they offer a convenient bridge for passing over from the scientific handling of the topic to the cruder notions in popular currency."[1]

There are such wide differences between individuals in their ability to memorise that it would be indeed surprising if some such faculty had not been presumed. Nevertheless when John Smith complains that he has a poor memory, he is liable to be asked several pertinent questions. La Rochefoucauld pointed out that "Everyone complains of the badness of his memory but nobody of his judgment", but it is here assumed that John Smith's complaint was not a gesture *de haut en bas* that memory is the prerequisite of those less endowed with intellect.

The first question to John Smith might be with regard to his definition: Do you include under memory "the facts of habit, of physical growth, of physiological functioning, of instinct, of heredity" according to the teaching of Ewald Hering, Samuel Butler and Richard Semon? Semon refers to all such phenomena as mnemic, and Sir Percy Nunn has developed a theory of considerable educational importance which may be concisely expressed: Memory is conscious mneme just as conation is conscious horme.[2] But probably John Smith would not wish to treat the matter so comprehensively, and would be content to define memory in the words of the dictionary as "the power of retaining and reproducing mental or sensory impressions". Accepting this definition, the second question would be: Is it poor for colours or sounds, for names or faces, for magazine articles or Latin vocabulary, for past occurrences or future engagements? It may be poor in some respects but surely not in all? In the words of Spearman: "When two kinds of memory resemble each other only in the bare fact of both involving retentivity, the correlation arising from this cause is little if at all above zero. In proportion as the likeness between them is augmented by resemblance of material—for instance, by both being sensory, or by both being verbal—the correlation becomes more and more marked."[3] If John Smith does per-

[1] C. Spearman, *The Nature of Intelligence*, p. 241.
[2] T. Percy Nunn, *Education: Its Data and First Principles*, pp. 19-22.
[3] C. Spearman, *The Abilities of Man*, p. 290.

form poorly in all the above indicated respects, it will become apparent in the sequel that what is seriously called into question is not his retentivity but his "g" (cf. Chapter 8). Such considerations indicate that memory is a complex term even if it is not taken in the wide sense of Hering, Butler or Semon. Even if it is agreed to define it in the words of the dictionary its complexity still causes differences of opinion as to its nature, both in spite of and in consequence of the manifold researches made. Spearman has pointed out that when anything is remembered it is not altogether a case of bare retentivity, for there are two very active further operations involved: We form an awareness of our own experience as something passing in time, and we fill up the gaps which have been made in this awareness by the ravages of forgetfulness. The first is constructive, the second reconstructive, but both involve relations and correlates and it is not therefore surprising that such a test as that of "logical memory" should correlate with "g".

Bare retentivity then is a much simpler concept than that of memory, yet it remains largely unknown and seldom appears in the index of a psychological textbook.

LAW OF RETENTIVITY

Spearman's Law of Retentivity states that "Cognitive events by occuring establish dispositions which facilitate their recurrence". The Law of Association is regarded as a special case, that is, cognitive events by accompanying each other establish dispositions to do so hereafter.

Two fundamental problems at once come up for consideration, whether retentivity is related to "g" and whether retentivity is a unitary function.

RELATION BETWEEN RETENTIVITY AND "G"

It is not here implied that definite views have been formulated but rather that statements are often made which would appear to assume a close and direct relation between retentivity and "intelligence", *e.g.* to define "intelligence" as the capacity to

learn, or to define it as the ability to establish mental bonds, or yet again to define it as the brain's ability to preserve traces or engrams of its experience.

Curiously enough, statements are also made which would lead the reader to assume that retentivity and "intelligence" are antagonistic. There is the observation of Aristotle that "persons who have good memories are apt to be slow of wit" and there is the statement by Rochefoucauld already quoted.

When the matter was put to an experimental test it was seen that neither view can be adopted. The experimenter's task is, however, complicated by the fact that in general both eduction, that is, "g" and reproduction, that is, retentivity, co-operate. He has, therefore, to pick out instances where the process at work is largely eductive and other instances where it is largely reproductive.

McCRAE'S INVESTIGATION.—The Stanford Binet tests were analysed and sorted according as they seemed to involve more essentially eduction or reproduction.[1] McCrae concluded that the criterion of the value of tests of general ability is how much they demand the eduction of relations and correlates and not how much they demand of retentivity. Retentivity as such would appear to be independent of "g".

DOES RETENTIVITY POSSESS FUNCTIONAL UNITY?

Turning now to the second fundamental problem, it can be stated at once that retentivity does not possess functional unity for in that case, there would exist a second universal factor alongside of "g". But such existence of two universal factors would be in contradiction with the tetrad difference criterion. (see Chapter 23). It would also contradict common observation. But there may be, nevertheless, functional unity or a group factor, covering a part of the field. In fact, until experiment proves the contrary, there may be several such group factors. It will be convenient now to consider if there is any evidence in support of certain group factors which have been posited.

[1] C. R. McCrae, *Some Testing of Physically Defective and of Mentally Defective Children*, Rep. of Brit. Assoc. 1925, p. 352, and cf. C. Spearman, *The Abilities of Man*, p. 275. A short summary is given on p. 88 of this book.

RECALL *v.* RECOGNITION (cf. *Expt.* 1 and *Expt.* 2, p. 43)

Does the person who recalls one material well also recall another material well? Does the person who recognises one material well also recognise another material well? Does the person who recalls one material well also recognise that material well? The answer in each case, according to Achilles,[1] is "Not necessarily". There was no evidence of a group factor of recall or a group factor of recognition.

Ang Lanfen Lee[2] later employed a similar technique to Achilles, namely, to test the retention of four kinds of material, pictures of familiar objects, simple nouns, geometrical forms and nonsense syllables by the method of recall and also by the method of recognition. The correlations between the various recall tests were on the whole fairly high. Also those between the recognition tests. But the correlations between recall and recognition were very low. Both Achilles and Lee, however, seem to have neglected to test the "g" of their subjects.

REPRODUCTION OF UNIQUE EXPERIENCES AND REPRODUCTION AFTER REPETITION

Another attempted division of memory is that sponsored by Bergson. The object in choosing the experiments which are described on pp. 44-47 was to give an opportunity to discuss the original theory of Bergson who holds that there are two forms of memory. "The memory of the lesson, which is remembered in the sense of learnt by heart, has *all* the marks of a habit. Like a habit it is acquired by the repetition of the same effort." "The memory of each several reading on the contrary . . . has *none* of the marks of a habit. . . . It is like an event in my life."[3] The past then survives under two distinct forms: first in motor mechanisms (habit), secondly in independent recollections (pure memory). The second then, according to Bergson, is memory *par excellence*, while the first is habit interpreted by memory rather than memory itself.

[1] E. M. Achilles, *op. cit. Archives of Psychology*, No. 44, 1920.

[2] A. L. Lee, "An Experimental Study of Retention and its Relation to Intelligence", *Psychol. Mon.* vol. 34, 1925.

[3] H. Bergson, *Matter and Memory*, Engl. trans. p. 89.

Such a theory is not easy to test, for Bergson's two factors, "habit" and "pure memory", may both be in operation in a given test. But it occurred to Dr. May Smith and Professor McDougall[1] experimentally to bring about experiences which approximate to one rather than to the other and thus to put Bergson's hypothesis to a test.

"If all memory is of the habit type, then all tests which involve to any extent retention ought to correlate with those in which motor habit plays an obvious part, or at least they should correlate as highly with such as with others. If on the other hand Professor Bergson's distinction is fundamental, then we should expect the tests concerned with retention to fall into two groups; those of the habit type should correlate together and those predominantly involving pure memory should correlate together; but success in the one group would not necessarily carry along with it success in the other, *e.g.* we might find that people markedly good at tasks of the one type are weak at those of the other. The tests chosen were intended to illustrate both modes of retention, *i.e.* reproduction of unique experiences and reproduction after repetition."[2]

Evidence was adduced by these two investigators in support of Bergson's distinction between habit and memory. Two of their tests, the learning of nonsense syllables and the fixing of the movements of the typewriter contained all the characteristics of habit formation. Two others, the Blots Test and the Picture Test approximated to "pure memory", but it should be noted that the Blots Test was a test of recognition while the Picture Test was a test of recall. The subjects were forty-one women students and each subject was tested individually.

When it is considered that Bergson's theory, if substantiated, would have such intimate bearing on educational methods, it is seen how desirable it is to carry out further work on the lines initiated by these two investigators.

To this end Sur[3] recently carried out an investigation with Standard VI. boys and girls in Leeds, using four group tests

[1] M. Smith and W. McDougall, "Some Experiments in Learning and Retention", *Brit. Journ. of Psychol.* vol. 10, Parts 2 and 3, 1920.

[2] M. Smith and W. McDougall, *op. cit.*

[3] R. K. Sur, *loc. cit.*

which were labelled provisionally "Pure Memory Tests" and four likewise "Habit Memory Tests". The "Pure Memory Tests" were: (1) Substance Memory for Prose Passages, tested by number of ideas reproduced.[1] (2) Memory for Stories, tested by a questionnaire. (3) Memory for Coloured Pictures. (4) Memory for Words, using thirty monosyllabic nouns, each noun being exposed once for one second. The "Habit Memory Tests" were: (1) Memory for Words, as in the last test, but the children were now also tested after four and after seven presentations. (2) English-Hindustani Vocabulary Test. Twenty pairs of nouns were presented, one pair at a time, then the Hindustani words were presented in a prearranged order and the children were asked to give the English equivalents and were tested after one, four and seven presentations. The average of the results after four and after seven presentations was taken as the score. (3) Symbol-Digit Substitution Test. (4) Modified Code Test. In addition the children were tested by the Tomlinson "West Riding" Tests of Mental Ability.

It was found that every one of the eight memory tests correlated appreciably with mental ability, but there was no evidence that tests taken from the so-called "pure memory group" correlated higher with mental ability than those taken from the so-called "habit group". Moreover, the results did not support Bergson's view that "pure" and "habit" memories can be contrasted as involving two different mental processes. This was evident from the table of inter-correlations and also from the table of partial correlations after eliminating the factor of mental ability. Consequently teachers who base their practice on Bergson's theory when dealing with memorisation are building on insecure foundations.

METHODS FOR TESTING MEMORY

(1) METHOD OF COMPLETE MASTERY.—This was devised by Ebbinghaus (cf. H. Ebbinghaus, *Memory* (1885), translated by H. A. Ruger and C. E. Bussenius in 1913). The series is repeated until S can repeat it without error. The score is the number of repetitions required.

[1] Cf. M. E. Bickersteth, "Mental Tests to Children of Various Ages", *Brit. Journ. of Psychol.* vol. 9, Part 1, 1917, p. 33.

(2) Saving Method.—A varying interval is allowed to elapse after the series has been learnt by the method of complete mastery. The number of repetitions to relearn is noted and compared with the original number. This was also devised by Ebbinghaus.

(3) Memory Span Method.—This was devised by Jacobs (*Mind*, 1887). It was further developed by Ebert and Meumann (*Archiv f. d. gesamte Psychol.* 1905). A series of gradually increasing length is presented and the score is the maximal length which can be reproduced.

(4) Method of Retained Members.—So named by Ebbinghaus, this was developed by Pohlmann. It consists in allowing S a given number of repetitions so that the series is incompletely learnt. The score is the number of items correctly reproduced (cf. A. Pohlmann, *Experim. Beiträge zur Lehre vom Gedächtniss*, 1906; also *Zeitsch. f. Psychol.* 44, 1907, 134-140).

(5) The Method of Right Associates.—A series of pairs is presented, the first unit of each pair being accented. The score is the percentage of correct responses when only the first unit of a pair is again presented. The time of response is sometimes also measured. The method was devised by Jost and developed by Müller and Pilzecker (cf. A. Jost, *Zeitsch. f. Psychol.* 14, 1897, 436-472; G. E. Müller and A. Pilzecker, *Zeitsch. f. Psychol.* Ergänzungsband, 1900).

(6) The Method of Prompting.—This was devised by Ephrussi (cf. *Zeitsch. f. Psychol.*, 1904, 37). S is allowed a given number of repetitions so that the series is incompletely learnt. The score is the number of times S must be prompted in order to be able to reproduce the series.

(7) The Reconstruction Method.—After the series has been presented in a definite order, S views the series again in a chance order and attempts to arrange them in the original order. The score is the number and size of the mistakes (cf. H. Münsterberg and J. Bigham, *Psychol. Rev.* vol. 1, 1904; E. Gamble, *Psychol. Mon.* No. 43, 1909).

(8) The Method of Recognition.—A limited number of items is presented. They are then presented again in conjunction with other items. The score is the percentage of correct

recognitions. The method was used by J. M. Baldwin and W. J. Shaw at Toronto in 1895 (cf. J. M. Baldwin, *The Mind*, p. 160), by W. G. Smith (*Brit. Journ. of Psychol.* vol. 1, 1905) and by V. Henri (*L'Année psychol.* 8, 1901).

(9) METHOD OF REPORT (Aussage).—The material is usually a picture which is presented for a definite time and S gives either (1) a free account of narrative or (2) replies to a series of questions, or may give (1) followed by (2). He may also be asked to underline those portions for whose correctness he would be willing to attest. If P is the possible number of items to be reported and $c(r)$ the number of items rightly reported and underlined, then, if $c(r) = P$, he would be an ideal witness. A full account of the method is given in G. M. Whipple's *Manual of Mental and Physical Tests*, vol. 2, pp. 17-42.

QUESTIONS

1. "*Memini etiam quae nolo; oblivisci non possum quae volo*" —I remember the very things I do not wish to; I cannot forget the things I wish to forget—(CICERO, *De Finibus*). "The forgetting of impressions and experiences shews the working of the tendency to ward off from memory that which is unpleasant" (FREUD, *Introductory Lectures on Psycho-Analysis*). Does your experience agree with Cicero's or with Freud's or with neither? (cf. a paper by C. Fox, and another by A. Wohlgemuth in *Brit. Journ. of Psychol.* vol. 13, Part 4, April 1923 and one by H. Cason, *Arch. of Psychol.* No. 134, 1932).
2. The name of a person, whose appearance can be recalled, has been forgotten. Sometimes it can be remembered by simply going through the alphabet. Why is this?
3. Point out the advantages and disadvantages of using nonsense syllables in experiments on memory.
4. Test the fidelity of report of a group of subjects and analyse qualitatively the errors which they make.
5. Compare a subject's fidelity of report when determined (*a*) by the narrative method, and (*b*) by the questionnaire method.

References

1. H. Ebbinghaus, *Memory, A Contribution to Experimental Psychology* (1885), trans. by H. A. Ruger and C. E. Bussenius (1913). Teachers' College, Columbia Univ., New York.
2. C. S. Myers, *Textbook of Experimental Psychology*, Part 1, chaps. 12 and 13, 1911.
3. G. M. Whipple, *Manual of Mental and Physical Tests*, Part 2, chap. 9, 1915.
4. C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition*, chap. 19, 1923.
5. B. Edgell, *Theories of Memory*, 1924.
6. C. Spearman, *The Abilities of Man*, chap. 16, 1927.
7. W. S. Hunter, chap. 15 in *The Foundations of Experimental Psychology*, 1929, ed. by C. Murchison.
8. F. C. Bartlett, *Remembering: a Study in Experimental and Social Psychology*, 1932.

CHAPTER 6

IMAGERY

SECTION 1: SENSORY AFTER-IMAGES

Apparatus.—Red, blue and black papers, one inch square, preferably of standard Hering colours, with a white thread gummed to each underneath a corner. A large sheet of standard white paper.

For ordinary class work the tests may be taken in broad daylight so long as the papers and Ss' eyes are shielded from direct sunlight. For more exact work the tests should be taken in the dark room, using a standard lamp, invisible to S, which illuminates the colours from a fixed distance.

Directions.—S fixates for 15 seconds the centre of the small square which is placed at the centre of the large sheet. E then jerks the small square away by means of the attached string, but S continues to fixate the place where the small square was held. S says "Come" when the first after-image appears, and "Gone" when it disappears. S says "Come" when the second after-image appears and "Gone" when it goes, and so on until no further changes occur. E, using the stop-watch, notes the times of these responses. S also describes the appearance of the images and the changes which occur.

It is recommended that these tests be taken with children.

Note particularly those children who have after-images of long duration and which tend to be continuous rather than to vanish periodically and reappear. Note also if any have after-images in the original colour and not in the complementary colour. Both according to E. R. Jaensch are symptoms of an eidetic faculty.[1]

SECTION 2 (A): SPONTANEOUS MEMORY IMAGES

Directions.—Each S is given a card and at a signal turns it

[1] E. R. Jaensch, *Eidetic Imagery*, 1930, pp. 5-7.

Visual Image of—

(1) the front door of your house
(2) a shilling
(3) a motor bus
(4) a tooth brush
(5) a watch

Auditory Image of—

(6) the horn of a motor car
(7) a crying child
(8) thunder
(9) a violin
(10) the bark of a dog

Olfactory Image of—

(11) coffee
(12) tobacco
(13) fresh paint
(14) cheese
(15) freshly cut grass

Gustatory Image of—

(16) chocolates
(17) an orange
(18) vinegar
(19) salt
(20) sugar

Tactual Image of—

(21) smooth glass
(22) sand paper
(23) velvet
(24) canvas
(25) wool

Kinaesthetic Image of—

(26) climbing stairs
(27) rowing
(28) shaking hands
(29) swimming in surf
(30) throwing a ball

Thermal (cold) Image of—

(31) ice in the hand
(32) cold wind
(33) cold shower bath
(34) stepping barefooted on cold linoleum
(35) ice cream in the mouth

Thermal (hot) Image of—

(36) a hot bath
(37) hot potato in mouth
(38) heat from a fire
(39) heat from a dish held in the hand
(40) heat from the sun's rays

Algesic Image of—

(41) the prick of a pin
(42) a burn
(43) pain in a tooth
(44) sore throat
(45) headache

Complex Image of—

(46) thirst
(47) hunger
(48) remaining under water
(49) sensations after a long day's walk
(50) well being (above par)

over and reads the question and solves it "mentally" without writing or tracing or using the card or any other object as an aid. S then gives an exact account of the method employed to arrive at the answer. The following questions are recommended, each appearing on a separate card:

(1) My house faces the street. If a boy passes my house in the morning, walking toward the rising sun, with my house at his right, which direction does my house face?

(2) Think of a triangle. From the centre draw a line to any of the corners. Also draw lines from the centre to the middle of each of the three sides. Into how many parts is the triangle now divided? How many of the parts are triangles?

(3) Think of a square standing vertically. Draw a horizontal line right across it dividing it into two equal parts. From the middle of the top side of the square draw lines to the bottom corners of the square. Into how many figures is the square now divided? What is the shape of the figures?

(4) Think of a square. From the middle of the top line draw lines to the lower corners. From the middle of the bottom line draw lines to the upper corners. Into how many parts is the square now divided?

SECTION 2 (B): VOLUNTARY MEMORY IMAGES

Read through the following list and select five which you think you can image as clear as any in the list. Give them a value of 6 and record this opposite their respective numbers. Then read the list again and score the clearness of the voluntarily aroused image in each case. Thus the value 6 would mean that the image was as clear as the standard. Then the values 5, 4, 3, 2, 1 would somewhat correspond to the "very good", "good", "moderate", "poor", "very poor" of ordinary marking, whilst the value 0 should be recorded when you fail to get an image.[1]

[1] This is a modification of Griffitts's method. Cf. C. H. Griffitts, "Individual Differences in Imagery", *Psychol. Mon.* vol. 37, 1927, p. 8.

The old method of scoring was to compare the clearness of each image with the clearness of perception. Here the method of Griffitts has been adapted for two reasons. Students frequently find it difficult to compare an image with a percept when the object is not there to be perceived, and some students overrate the clearness of all their images.[1]

Which kind of voluntary imagery ranks highest in clearness?

Compare the voluntary imagery of men and women.

Calculate the correlation between the clearness of visual and of auditory imagery, also that between the clearness of visual and kinaesthetic imagery.

SECTION 3: EIDETIC IMAGERY

APPARATUS.—The Münchener Bilderbogen "Was alles am Morgen geschieht", published by Braun und Schneider, Munich. One of these pictures, about the size of a postcard, to be presented at a time.

Several other types of pictures about the same size which would appeal to children and which must have meaning as a whole should be available.

A homogeneous dark gray background corresponding to 50° white and 310° cloth-black on the colour disc.[2]

DIRECTIONS.—This test should be taken with children. One of the pictures is placed on the background at normal reading distance. Without fixating a particular spot, S looks at the picture as a whole for 15 seconds so as to get the general meaning. He then reports what he sees on the background after the picture has been removed.

Before undertaking this test it is advisable for the child to take one of the tests of Section 1 in order that he may realise that something is *seen* on the white background after the square has been removed although there is no object actually present. He will then be better able to answer whether he *sees* anything in the present test after the picture has been removed. E should endeavour to get the child to give an exact account of what he

[1] C. H. Griffitts, *loc. cit.*
[2] E. R. Jaensch, *loc. cit.*

sees, by suitable questions as to the size, colour, clearness, localisation, movement and duration of the images.

E decides how many children should be tested, *e.g.* he may attempt to select six children possessing tetanoid characteristics and six possessing basedowoid characteristics (cf. the discussion on this chapter).

DISCUSSION

The study of imagery would appear to be the most baffling in the whole realm of psychology. The behaviourist usually ignores the topic since he regards it as unworthy of discussion owing to its speculative and unreliable nature. Such an attitude is not surprising, for as Köhler points out: "One psychologist claims to have them (images) in numbers, many of them almost as lively and concrete as perceptions. Others tell us that there are no real images at all in direct experience."[1]

It is not difficult to account for this lack of unanimity. Individuals differ so much with respect to the vividness of their images, whether spontaneous or voluntary, that the drawing of general conclusions is a very hazardous undertaking. Moreover, there are peculiarly insidious pitfalls in experimentation on imagery:

(1) There is the necessity to differentiate between voluntary and spontaneous imagery. It does not in the least follow that the individual who gets vivid voluntary images would experience such spontaneously in his intellectual activity.[2]

(2) If material is presented through a particular sense with a view to the subsequent testing of imagery, it does not follow that those who excel in such a test possess vivid imagery in that sense; thus if material is presented visually, a weak visualiser might have very vivid auditory images which might conceivably enable him to do as well in the test as a strong visualiser. Thus Diamanti, whose ability for retaining numbers was phenomenal, used visual imagery, whilst Inaudi performed similar feats with the aid of auditory images.

(3) There has been a tendency to force individuals into types

[1] W. Köhler, *Gestalt Psychology*, Amer. ed., 1929, p. 9 (Engl. ed. 1930).

[2] Although Dr. Carey found appreciable correlation between spontaneous and voluntary imagery it was not high enough to be prognostic. Cf. *Brit. Journ. of Psychol.* vol. 7, 1914-15, p. 466.

according to the kinds of images which dominate their thinking, and also to assume further that the individual with indistinct visual imagery must, of necessity, have clear and distinct imagery of some other kind. Without denying that there are some individuals who may be termed visiles, audiles, motiles or what not, there is yet much force in the words of Thorndike: "Instead of a few 'pure' types or many 'mixed' types, there is one type—mediocrity".[1]

(4) Psychologists have now to re-examine imagery in view of the theories and conclusions of the Marburg school regarding eidetic imagery. The best introduction to the subject is found in a book entitled *Eidetic Imagery* by Professor E. R. Jaensch, the leader of the Marburg school. It has been demonstrated that many children and even some adults, after looking at a suitable picture in accordance with the directions given in Section 3 of this chapter, are able to see an image of the picture on the projection mat after the picture has been removed. The image may come up immediately or after an interval. Sometimes it is in the original colours, sometimes in the complementary colours, and sometimes in gray. Jaensch regards eidetic images as phenomena that take up an intermediate position between sensations and images. When imagination has little play they approximate to sensory after-images. When imagination is at a maximum they can be regarded as memory images which are literally visible. The percentage of children who can see these images is said to vary with age, race, locality and methods of instruction.[2] But the investigation of Teasdale has shown that the discrepancies between the results of the various investigators is partly due to their adopting different criteria as to what constitutes an eidetic image. Roessler[3] found the eidetic disposition at its maximum amongst children of six years, and these were the youngest children tested. Jaensch gives this view his support. Others[4] place the maximum at

[1] E. L. Thorndike, *Educational Psychology*, vol. 3, 1914, p. 374.

[2] Cf. Margarete Zimmermann, *Eidetik und Schulunterricht*, 1931, p. 37, and E. R. Jaensch, *Eidetic Imagery*, 1930, p. 9.

[3] F. Roessler, *Beiheft* 43 *der Zeitschr. f. angew. Psychol.* 1928.

[4] Cf. H. Zeman, *Zeitschr. f. Psychol.* 96, 1925; S. Fischer and H. Hirschberg, *Zeitschr. f. d. ges. Neur. u. Psychiat.* 88, 1924.

various points between twelve and fiteen years. Now Teasdale's results indicated a definite correlation between the duration of the eidetic image and the richness of its details. In the system of marking finally adopted it was possible to gain additional credit according to the duration. Teasdale[1] found that if only subjects with exceptionally good eidetic images are considered, then the ability decreases with age. If those with good or moderate images are also included, then the ability rises to a maximum at eleven to twelve years. If all who get images are included then the ability increases with age.

The study of eidetics by the Marburg psychologists is not limited to mere imagery and theories of perceptual development, but is also made the basis for a typological study of human personality. Pronounced cases of eidetic disposition are said to fall into two varieties: the T, or *tetanoid* form, where the images approximate to the sensory after-images, "the eyes being small, deep-set, comparatively lifeless, without lustre, with no 'soulful' expression". Pathological cases were said to benefit from feeding with calcium. On the other hand is the B, or *basedowoid* form, where the images approximate to memory images, the eyes being large and "soulful", the body gracefully built, the skin soft and silky with a low resistance to electric currents. It was further held that in the majority of youthful eidetic subjects both components are present, the pure B type being relatively frequent, the pure T type rare.

Far-reaching claims as to the importance and significance of the typological method in education, in medicine, and in biology have been made by Jaensch. "Our procedure in eidetics, in the psychology of childhood and in wide fields of general psychology, consists in attempting to clear up the broad mass of useful cases by starting with a few special cases to guide us. These special cases have certain characteristics in an exaggerated form, which stick out above the mass, as it were, and are easily discernible. In that way the apparently chaotic and diffuse picture presented by the mass can be made clear."[2]

[1] H. Teasdale, "The Prevalence of Eidetic Imagery among Schoolchildren and its Educational Significance", M.Ed. Thesis, Univ. of Leeds, 1932.

[2] E. R. Jaensch, *op. cit.* p. 60.

(5) In all experiments on imagery, whether sensory after-images, eidetic images, or memory images it is essential to repeat the tests a sufficient number of times. Thus the present writer found that the tests on after-images described in Section 1 have not much validity unless repeated on several days. But it was found that the after-images of certain children tended to be continuous for a fairly long time in comparison with those of certain others which tended to come and go periodically. It will be seen that Jaensch attaches importance to this difference in singling out some of his eidetic cases. It is a question which seems to deserve further study. The present writer, as a result of preliminary experiments, suspected a relation between perseveration and the duration of the after-images.

This is not the place to discuss debatable points but rather to indicate those on which the majority of psychologists would agree. Most would assert that the great majority of individuals have imagery of various kinds, of sights, sounds, movements and the like and would find it easy to distinguish between an image and a percept. It is therefore rather extraordinary that difficulty is experienced in stating what exactly constitutes the difference between a percept and an image. It is true that an image is usually less clear, that the percept depends on our movements while the image depends on our will, and usually the percept has a reality which at any rate is not possessed by the memory image. But that is not the whole story and most would probably agree with the weighty testimony of Wohlgemuth: "That these memory-images are faint reproductions of the sensations, *i.e.* weak sensations, sensations of low intensity, is decidedly not true in my case; they are an experience *sui generis*".[1]

In agreement at least would be the experience of the present writer.

He possesses visual images of Professor and Mrs. C., his companions at Lake Mendota, faint auditory images of their voices, a visual image of the lake, a thermal image of its waters which seems a blend of invigorating coolness and pleasant warmth, a visual image of the vegetation of its bed as seen in the peculiar light

[1] A. Wohlgemuth, *Critical Examination of Psycho-Analysis*, 1923, p. 25.

of subaqueous vision. Each of these reproductive or memory images differs from a weak sensation, but even more so does the following image which sometimes arises: a visual image of the narrow neck of land with Indians transporting their canoes over it in order to shorten their voyage. Such an image of constructive imagination is obviously not a copy of a sensation, as Indians have probably not made such use of this neck for several generations.

IMAGERY AND LITERARY APPRECIATION

There are at least three studies bearing directly on the function of images in literary appreciation, those of Wheeler,[1] Valentine[2] and Mason.[3] Each advocates the cultivation of the imagery of children. Wheeler used her "image-formation method", but by this is not implied that children should make an effort to obtain relevant images in order to appreciate a poem, but that a highly trained teacher should read a new poem with expression and without disturbing its rhythm. The child then forms the relevant images and increases his appreciation. Wheeler criticises methods which are too prevalent: the method of exposition, the method of silent reading and the reading aloud by the pupil, and the method of repetition.

Valentine's study emphasises the caution which must be exercised before drawing conclusions owing to the decided individual differences exhibited.

Mason also advocates the use of the image-formation method as the most successful means of attaining literary appreciation, but finds that there is a minority, chiefly weak imagers, for which the method is definitely not the best. He pleads for the lengthening of the period in life during which the formation of images is possible, by a more deliberate acknowledgment of the importance of imagery in school. This view may be compared with Jaensch's claim that his surveys have shown that the eidetic phase is preserved longer than usual in children who are educated by methods adapted to the mentality

[1] O. A. Wheeler, "An Analysis of Literary Appreciation", *Brit. Journ. of Psychol.* vol. 13, Part 3, 1923.

[2] C. W. Valentine, "The Function of Images in the Appreciation of Poetry", *ibid.* vol. 14, Part 2, 1923.

[3] J. E. Mason, "Imagery and the Teaching of English", M.Ed. Thesis, Univ. of Leeds, 1928.

of youth, *e.g.* the methods of the modern Arbeitsschule in Germany.

QUESTIONS

1. Do you get many images in reading fiction or poetry? Do you find a tendency for a particular kind of imagery to be more vivid than the rest?
2. If visual images are more vivid and numerous than other kinds of imagery when you read poetry, does it follow that you belong to a visual type? If not, why not?
3. Do you use imagery to a large extent in solving geometrical problems?
4. Discuss the bearing of the psychological study of imagery on practical teaching in the elementary school.

References

1. C. Spearman, *The Nature of Intelligence* chap. 12.
2. S. I. Franz, "After-images", *Psychol. Mon.* vol. 3, 1899.
3. G. H. Betts, "The Distribution and Functions of Mental Imagery", Teachers' College, Columbia Univ., *Contributions to Education*, 1909, No. 26.
4. C. H. Griffitts, "Individual Differences in Imagery", *Psychol. Mon.* vol. 37, 1927.
5. E. R. Jaensch, *Eidetic Imagery*, 1930.
6. F. Galton, *Inquiries into the Human Faculties*, 1883.
7. N. Carey, "Factors in the Mental Processes of School Children. I. Visual and Auditory Imagery", *Brit. Journ. of Psychol.* vol. 7, Part 4, 1915.
8. G. Dawes Hicks, "On the Nature of Images", *ibid.* vol. 15, Part 2, 1924.
9. Symposium on "The Relevance of Visual Imagery to the Process of Thinking", by T. H. Pear, F. Aveling and F. C. Bartlett., *ibid.* vol. 18, Part 1, 1927.
10. B. Edgell, "Images", *Aristot. Soc. Proceedings*, 1930-31.
11. H. Teasdale, "A Quantitative Study of Eidetic Imagery", *Brit. Journ. of Educ. Psychol.* vol. 4. 1934.

CHAPTER 7

"TRIAL AND ERROR" LEARNING

MATERIALS.—Metronome beating seconds. Six copies of Fig. 12 to be prepared by each S. Pencils. Drawing-pins. Mirrors. Cardboard covers with suitable supports.

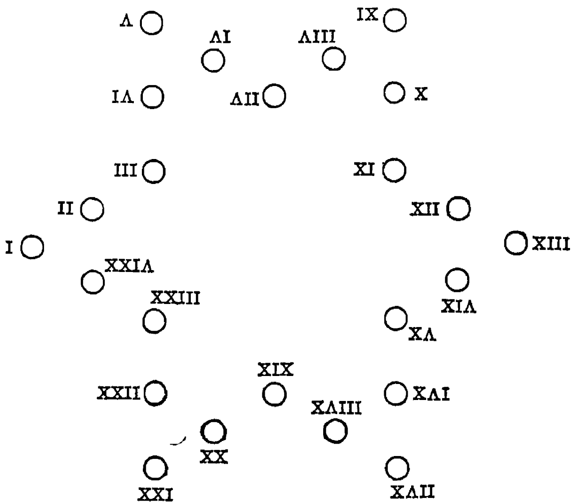

FIG. 12.

INSTRUCTIONS.—Students should work in pairs, each being E and S alternately until each has completed at least six attempts. Fig. 12 is fixed with drawing-pins squarely in front of S by E, who should also arrange the cardboard cover so that it will cut off S's view of the diagram but will allow him to see it in the mirror and will not interfere with his hand in drawing. The whole class should then be given the following instructions:[1]

"When E has placed the point of S's pencil at circle No. 1,

[1] Modified from M. Gopalaswami, "'Intelligence' in Motor Learning", *Brit. Journ. of Psychol.* vol. 14, Part 3, January 1924.

S should try to trace the pattern but be very careful to do so according to the following instructions.

(1) Only straight lines to be drawn.

(2) Make a line and only one line for every stroke of the metronome which beats seconds. Move in jerks, stopping after every stroke.

(3) If you go the wrong way, correct from where you are. Do not go back to the starting point. Also do not correct at once, but wait for the next stroke of the metronome, and then make the correcting movement.

(4) You must arrive inside a ring at the end of your stroke before proceeding to the next. Thus if you fail to stop inside a ring, you must come back to it.

(5) Always hold the pencil in a vertical position and employ arm movements, thus avoiding movements of the fingers or wrist only."

It is important to insist on the observance of the instructions, as uniformity of method is essential to get significant results. By the above procedure time and error measurements are reduced to one category, namely, that of the number of movements to traverse the pattern.

DISCUSSION

When the learned man errs he errs with a learned error.

Arabic Proverb

The credit for focussing attention on the analysis of the so-called "trial and error learning" is largely due to Thorndike. It is important to realise that his classic studies in animal psychology were completed before he embarked on his epoch-making researches in educational psychology. Thorndike's pioneer experiments undoubtedly developed comparative psychology, a science which now attracts so many workers, especially in America. Moreover, their influence is manifest in most educational textbooks which attempt to deal with the laws of learning.

Thorndike found that a hungry cat placed in a box will succeed in gradually decreasing the time needed to undo the fastenings and reach the food placed outside, if given repeated

trials. It is true that the improvement is not very regular, but in the end the time becomes fairly constant, and the animal has "learned" the operation of escaping from the box and reaching the food. No one has questioned the facts, but what of their interpretation? How did the animal "learn"? Obviously a difficult question, and it is not surprising that several theories have been propounded. In order to get corresponding data from human subjects the nature of the task would have to be more difficult, *e.g.* mirror writing or the solving of complicated ring puzzles. In general such learning curves obtained with human subjects are similar to those obtained by Thorndike with cats, except perhaps that with humans sudden drops are occasionally found corresponding to those points at which some underlying principle has become manifest. Koffka,[1] however, points out that these sudden drops are also exhibited in some of Thorndike's curves and argues that Thorndike has not sufficiently credited the animal for its insight while unduly stressing its stupidity.

Space does not admit a discussion of all the laws of learning which Thorndike formulated on the basis of his early experiments, but the Law of Effect and the Law of Use have figured so prominently in the theory and practice of teaching that it is important to become acquainted with Thorndike's more recent investigations with regard to them.

THE LAW OF EFFECT (OR LAW OF SATISFACTION AND ANNOYANCE)

Thorndike's original and too simple formulation may be summarised thus: the response which leads to satisfaction becomes facilitated and that which leads to annoyance hindered. Probably this is the usual meaning given to the law by school teachers. But Thorndike no longer believes that the strengthening influence of a connection by satisfying consequences is paralleled by the weakening influence of annoying consequences. He now holds that the former are more universal, inevitable, and direct; the latter more specialized, contingent upon what the annoyer in question makes the subject

[1] K. Koffka, *The Growth of the Mind*, 1924, p. 164.

do, and indirect.[1] In spite of various criticisms Thorndike is more convinced than ever that his Law of Effect is *the* fundamental law of learning.

THE LAW OF USE (OR LAW OF EXERCISE)

The original formulation meant, in brief, that the response which frequently follows a given situation becomes facilitated. But Thorndike now minimises the importance of mere frequency and holds that "If a certain state of affairs acts upon a man a thousand times a week for a year, he will, so far as the mere repetition of that state of affairs is concerned, probably respond no better the last week than the first".[2]

Thorndike still needs the Law of Use, for he believes in the potency of repetition when "belongingness" is present, but not when it is absent.[3] Thus in the words "Fridtjof Nansen drifted northwards for two winters. He then left the ship", there is much belongingness attached to "Fridtjof" and "Nansen" but very little to "winters" and "He".

One of Thorndike's recent experiments[4] which, he thinks, supports the Law of Effect was to present on a fixed background in a random order fifty strips of paper, two of every length between three and twenty-seven centimetres. Each subject had before him a strip ten centimetres long and known to be ten, and estimated the length of each strip in integral numbers. Two groups of similar general intelligence and initial ability were selected and the method of procedure for one group was to place a strip before the subject. As soon as he announced his estimate, the strip was withdrawn from view and the experimenter said "right" or "wrong" according as the estimate was right or wrong, No statement was made of the amount or direction of the error. After a number of such presentations of the fifty strips (usually seven, two or three a day for three or four days) the series was presented with no aid. The effect of calling "right" and "wrong" was to produce a

[1] E. L. Thorndike, *The Fundamentals of Learning*, 1932, p. 276.
[2] *Ibid.* p. 62.
[3] *Ibid.* p. 66.
[4] E. L. Thorndike, "The Law of Effect", *Amer. Journ. of Psychol.* vol. 39, 1927, p. 212.

considerable reduction of error, whereas the other group after equal practice but without getting the announcements of "right" and "wrong" did not improve in their estimates of the lengths.

It might be worth while to repeat the experiment in order to see if the reduction in error due to the announcements correlated with intelligence.

The inadequacy of the Law of Use in the sense that mere frequency would produce learning is shown in an ingenious experiment by Peterson.[1] His subjects learned to trace an imaginary maze although the conditions prescribed ensured that repetition of response would have hindered learning. Many writers on the psychology of learning show a tendency, when psychological explanations are lacking, to formulate a physiological hypothesis to account for the facts. In the course of time this hypothesis is itself regarded as a fact. Thus learning was said to be due to the repeated passage of a neural impulse through a synapse thereby reducing the synaptic resistance. But Lashley[2] states that "no direct evidence for changes in synaptic resistance has ever been obtained. The hypothesis is not based upon neurological data but is merely a restatement of the observed fact that increased efficiency follows repeated performance." In his experiments with rats Lashley proved that synapses which were functioning for the first time in a given situation did so perfectly.

The importance of the mirror drawing experiment is that it furnishes a means of studying what is commonly called learning by "trial and error". When S has attempted the task six times, let him try to analyse all the errors which he has made. This analysis will prove somewhat difficult to the beginner but his task may be considerably simplified if Spearman's principles are applied. For a detailed account of the application of these principles to mirror drawing the reader is referred to Professor Gopalaswami's important paper. It is there shown how the

[1] J. Peterson, "The Backward Elimination of Errors in Mental Maze Learning", *Journ. of Exp. Psychol.* vol. 3, 1920, p. 257, and "Learning when Frequency and Recency Factors and Negative", *ibid.* vol. 5, 1922, p. 270.

[2] K. S. Lashley, "Studies of Cerebral Function in Learning", *Psychol. Rev.* vol. 31, 1924, p. 369.

mistakes made in the foregoing experiment may be arranged in two levels, higher and lower respectively:

I. *Lower Level*—	II. *Higher Level*—
(1) habit	(5) correlate-finding, based on misapplying the relation of inversion
(2) "wild"	
(3) inhibition	(6) correlate-finding, based on misapplying the angular correction
(4) withdrawal	

Thus a habit error would be to move from I to X instead of from I to II and is therefore an instance of the process of reproduction, namely, that the presenting of any character reproduces some other character previously associated with it.[1]

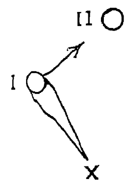

FIG. 13.

A "wild" error—very infrequent—seemed to occur from an impulse of despair to move anywhere.

By inhibition is meant the failure to move at all.

A withdrawal, such as from X to I, is a common reaction after a wrong one. It can be explained by supposing that the previous wrong movement was perceived to be such (Spearman's Second Principle) and that then the relation of withdrawal is reproduced because it is a habit to withdraw in such circumstances (Process of Reproduction). Finally, this relation together with the perceived present position at X jointly evoke the actual movement of withdrawal (Spearman's Third Principle).

The movement depicted in the next figure is explained as a case of misapplying the relation of inversion. A memory is present of the habit error already committed, which had been perceived (Spearman's Second Principle) to consist in the desired movement being something less than a right angle anti-clockwise of the movement actually executed. This anti-clockwise relation together with the present perceived position of the pencil evoke conjointly, as the correlate, such an impulse as would normally produce the

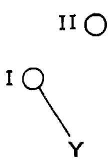

FIG. 14.

[1] C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition*, p. 137.

appropriate anti-clockwise correction (Spearman's Third Principle). But owing to the mirror the change of movement actually effected, namely from I to Y, has the reverse effect of a shift clockwise.

Finally, a movement such as that from I to Z in the figure, is a case of misapplying the angular correction. The subject soon becomes aware that the mirror engenders a vertical inversion (Spearman's Second Principle); accordingly, since he sees that his pencil should move slanting upwards in the mirror-space, he makes actually the movement that would in ordinary space take it slanting downwards. This constitutes an instance of applying Spearman's Third Principle. But it is executed inaccurately, the allowance for slant being insufficient.

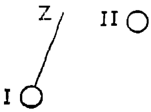

Fig. 15.

It is, however, necessary to remember that although the rapidity and ease with which the task is accomplished may be taken as an index of learning capacity, yet it is possible that the factor of perseveration (cf. Chapter 18) may handicap some individuals in so far as the new connections to be established are, in their case, more powerfully interfered with by firmly established habits. The task thus differs in one important respect from typewriting, morse-signalling or any task where it is not necessary to inhibit habitual reactions to such a large extent.

Snoddy concluded that a recess period in the early stages of learning makes possible the attainment of accuracy; and repetitions without recess in the later stages of the learning make possible the attainment of speed. By a "recess period" is meant the period of time which the experimenter interpolates between successive efforts of the learner, each effort being called a unit of practice.

There is even yet but little evidence as to the length of the optimal recess periods in various scholastic activities, and as to the most economical distributions of repetitions.

Book,[1] as a result of his investigation of typewriting, concluded that practices without recess lead to ruts and limit the

[1] W. F. Book, *The Psychology of Skill*, 1908.

spontaneity of movement. Gopalaswami,[1] in a special form of the mirror drawing experiment, found that the distributed method produced a much larger proportion of responses of the higher levels; whereas the massed method caused a large increase in the responses of lower level, especially those of mere "habit".

Until recently it was customary to say that improvement in mirror drawing occurred through trial and error and that reasoning as such had little or no part. This attitude is characteristic of all "mechanical" theories which involve the notion of "random" or "chance" movements. Opposed to such views is the "rational" theory of Professor Spearman. According to this theory "rational" or "intelligent" movement plays an important part in mirror drawing. Moreover, "intelligence", according to this theory may be defined as the power of noegenesis, if it is agreed to call the processes of educing relations and correlates noegenetic processes. It is in favour of the theory that the efficiency of children in the mirror drawing experiment correlated highly with their intelligence as estimated by their teachers. Moreover, an analysis of the introspections of thoroughly trained psychologists led to the same conclusion. Finally the theory opens up new possibilities for training in manual skill. For when an operation has been analysed into its ultimate processes and then the learner initiated into their nature, his skill will become facilitated.

QUESTIONS

1. Do you find it possible to arrange your wrong reactions in the last experiment in the six classes mentioned?
2. Show from your records how far (*a*) associative reproduction, (*b*) the educing of relations, (*c*) the educing of correlates, dominate the learning.
3. If "trial and error" operations as well as "intelligent" operations involve the educing of relations and correlates, how can the large differences between the two kinds of learning be explained?

[1] M. Gopalaswami, *loc. cit.* pp. 281, 285.

4. Having analysed the operation of mirror drawing into its ultimate processes, show how the progress of a learner could be facilitated.
5. Then show how you would determine experimentally the degree of facilitation.
6. "Learning is often said to take place either by practice (trial and error), by imitation, or by some form of ideational control (instruction, reasoning, etc.). In the mirror drawing test the conditions preclude the use of imitation and there is but relatively little opportunity to employ ideational control; whatever improvement appears is due primarily to a process of trial and error" (WHIPPLE, *Manual of Mental and Physical Tests*, vol. 2, p. 119). Discuss the above statement with respect to the learning of mirror drawing.

REFERENCES

1. M. Gopalaswami, "'Intelligence' in Motor Learning", *Brit. Journ. of Psychol.* vol. 14, Part 3, January 1924.
2. G. S. Snoddy, "An Experimental Analysis of Trial and Error Learning in the Human Subject", *Psychol. Mon.* No. 124, 1920.
3. G. M. Whipple, *Manual of Mental and Physical Tests*, vol. 2, p. 119.
4. E. L. Thorndike, *The Fundamentals of Learning*, 1932.

CHAPTER 8

THE "G" FACTOR AND TESTS OF "INTELLIGENCE"

MATERIALS.—Tomlinson's "West Riding" Tests of Mental Ability.[1] Price 3s. 6d. per dozen. Examiner's Manual, 6d. Published by Hodder and Stoughton. Four dozen copies of Set Y and also of Set Z should be available so that any class in the elementary school from Standard IV. upwards may be tested. The copies may be used repeatedly as the answers are written on other sheets.

DIRECTIONS.—E should arrange with the headmaster of a school for the testing, say, of the pupils in Standard VII., and for obtaining their order of merit according to their school marks and also according to their teacher's estimate of their general mental ability.

The students should witness the administration of Set Y by E and should endeavour to form a critical estimate of the procedure. These estimates may then be discussed in class. With the aid of the Examiner's Manual, each student should score the answers of at least one pupil. The correlation between the order according to school marks and that obtained from Set Y should be calculated. Likewise, the correlation between the order according to the teacher's estimate and that obtained from Set Y. If desired, the same pupils may be tested the following week, using Set Z, and the corresponding correlations calculated. In addition, the correlation between the two school orders may be calculated. (For particulars cf. Chapter 23.) The conclusions drawn should be discussed. By reference to the Examiner's Manual the Intelligence Quotient of each pupil is easily obtained. As a check two pupils may be picked out and each may be given an individual test, *e.g.* Burt's London Revision of the Binet-Simon Scale.[2] This serves two purposes. It acts as a

[1] Cf. T. P. Tomlinson, "A Group Scale of Mental Ability", M.Ed. Thesis, Univ. of Leeds, 1923.

[2] C. Burt, *Mental and Scholastic Tests*, 1921 (3rd impression, 1927).

demonstration of individual testing, even though the child is tested in front of an audience, which is a condition always to be avoided in exact individual testing. In addition, this procedure shows roughly the degree of agreement between the Intelligence Quotient supplied by the Tomlinson Scale and that obtained from a Binet-Simon Scale. Further, if the position of a child in the "West Riding" Tests differs markedly from his position as given by the school, the Binet-Simon Tests may furnish additional evidence to account for the discrepancy. Finally an attempt should be made to analyse the nature of the cognitive relations which have to be educed in each type of test employed.

DISCUSSION

> So much for the master's motto, "Every man in his place". Next for the labourer's motto, "Every man his chance". Let us mend that for them a little, and say, "Every man his certainty"—certainty, that if he does well, he will be honoured, and aided, and advanced in such a degree as may be fitting for his faculty and consistent with his peace. . . . It is the law of good economy to make the best of everything. How much more to make the best of every creature!
>
> J. RUSKIN, *The Crown of Wild Olive*
> IV. The Future of England

> In some sense and in some effectual degree, there is in every man the material of good work in the world; in every man, not only in those who are brilliant, not only in those who are quick, but in those who are stolid, and even in those who are dull.
>
> W. E. GLADSTONE

(Cf. C. Spearman, *The Abilities of Man*, p. 221.)

What is intelligence? On what principles are the component tests in an intelligence scale selected and assembled? Why are the particular time limits adopted for each test? Are there not individuals of high intelligence but slow? Would not such be penalised by the time limits adopted? Is it not possible to "coach" pupils for such tests? What are the practical uses of intelligence scales?

These are samples of the avalanche of questions which may arise at the ensuing discussion. Comprehensive answers would demand a volume. In this book a brief discussion of some important aspects must suffice.

The whole subject of intelligence testing becomes greatly

clarified if it be realised that what should be tested is the individual's efficiency or power in the operations which are covered by the three qualitative laws or principles already stated, namely, the excellence of his self-awareness and of his eductions of relations and correlates, always assuming that affective and conative factors are being kept under control. When it is further remembered that the analysis of self-awareness has as yet hardly figured in the testing of intelligence, it will be apparent why every successful test of intelligence so far devised gives the testee an opportunity to educe relations and correlates of various kinds. The complete list of relations according to Spearman[1] is as follows: Evidence, Likeness, Conjunction, Space, Time, Psychological, Identity, Attribution, Causation, Constitution.

Simple illustrations would be:

RELATION OF EVIDENCE.—John is taller than William, and William is much taller than Bob. Compare John with Bob.

RELATION OF LIKENESS.—Draw a line under the word which means the same or nearly the same as the word "command": obey, surrender, order, fire, run.

RELATION OF CONJUNCTION.—63 + 14 = ?

RELATION OF SPACE.—Cancel every *a*, every *n*, every *o* and every *s* on a printed page.

RELATION OF TIME.—Cf. Seashore's Sense of Time Test, Chapter 15.

PSYCHOLOGICAL RELATION.—This is involved whenever we interpret the thoughts, desires or actions of other people by analogy to ourselves. The best known test of this is Binet's Interpretation of Pictures Test.[2]

RELATION OF IDENTITY.—The Thurstone Hand Test.[3] Pictures of the right hand and of the left hand, in various positions, follow each other in haphazard order, and S has to underline each picture of the right.

RELATION OF ATTRIBUTION.—Tiger is to Fierce as Lamb is to?

[1] C. Spearman, *The Abilities of Man*, pp. 168-184.

[2] Cf. C. Burt, *Mental and Scholastic Tests*, p. 26.

[3] Obtainable from Messrs. C. H. Stoelting, 424 North Homan Avenue, Chicago.

RELATION OF CAUSALITY.—Prisoner is to Jail as Water is to Prison, Drink, Tap, Bucket. Which of the last four words is the answer?

THE RELATION OF CONSTITUTION.—This relation is of a secondary nature in that it springs from one of the other nine classes of relation. Thus the triangularity and the blackness together with their inter-relation constitute the black triangle below:

FIG. 16.

Further discussion is beyond the scope of this book, but it would appear that the relation of constitution must be considered in studying various Austrian and German doctrines regarding "shapes", "forms", "complexes", "structures" and "wholes". It will also elucidate Wundt's principle of creative resultants[1] as exemplified by his statement that "the product derived from any number of elements is more than the mere sum of the elements."

Space does not allow a fuller discussion of these ten relations, but there remains one fundamental question which must be faced: Is ability in educing one kind of relation any evidence concerning ability in educing another kind? Let us assume that, for each of the relations mentioned above, many different tests are available. Then at least seven hypotheses may be examined.

(1) There may be a general factor "g" extending throughout all the tests; and over and above this a group factor f_1 extending throughout all the tests involving the first class of relations, or throughout a number of them; similarly a group factor f_2 extending throughout all the tests involving the second class, or throughout a number of them, and so on for the other classes; and finally a large number of specific factors, the range of each being so narrow that it only plays a part in one particular test. Thus in a particular test a_1 involving the first class of relations, an individual possessing "g" in large measure would tend to do well, also if he possessed a large f_1, and also if he possessed a large sa_1, the factor specific to this test. His resultant performance would thus depend on the relative importance of these

[1] W. Wundt, *Outlines of Psychology*, trans. by C. H. Judd, 3rd ed. 1907, p. 369.

factors in the test. If this hypothesis were substantiated, intelligence testing would certainly be a complicated task.

(2) A simpler hypothesis would be to assert that there is only "g" and that the person who excelled in one kind of relation finding would excel in every other kind.

(3) A third would be to assert that there are only *f*'s, the group factors, in operation.

(4) A fourth that there are only *s*'s in operation, and that ability in one test has no relation to ability in another.

(5) A fifth might posit both *f*'s and *s*'s but no "g".

(6) A sixth might conceivably assume the presence of "g" and *f*'s but no *s*'s.

(7) Finally it is open to assume the presence of "g" and the *s*'s.

In order to decide between these hypotheses the tetrad difference criterion of Professor Spearman is a potent help (cf. Chapter 23). For if tests of different abilities are given to a large number of subjects, it can be established that whenever the tetrad equation holds throughout the table of correlation, and *only* when it does so, then every individual measurement of each ability can be divided into two independent factors, namely, a general factor "g" which varies from individual to individual but remains the same for any one individual with respect of all the correlated abilities, and a specific factor *s* which not only varies from individual to individual, but also for any one individual from each ability to another. This is the famous Two-Factor Theory which was enunciated in 1904 in a paper[1] which for two reasons is epoch-making: (1) The method of correlation was first successfully applied in psychology. (2) The doctrine enunicated has caused more discussion than any other in the present century. It is significant that although Spearman had published papers typical of the work coming from the laboratory of Wundt, who was then at the zenith of his career, yet in this paper both the method and the doctrine are new and independent of the Wundtian tradition. As often occurs in the progress of science it was several years before the

[1] C. Spearman, "General Intelligence Objectively Determined and Measured", *Amer. Journ. of Psychol.* vol. 15, 1904, pp. 201-292.

full significance of the paper was realised. Nevertheless the year 1904 marks the birth of a new school of thought.

It may at once be stated that when these various classes of relations were subjected to the tetrad difference criterion by Spearman and his pupils, the general result is that the "g" factor is common to all kinds of relations and it would be difficult to prove that any class of relations is more characteristic of "g" than another. Further, the number of group factors of wide extent in any particular class of relations is apparently very small. There is some evidence of such a factor somewhere in the domain of relations of evidence, namely, a logical factor;[1] another in the domain of spatial relations akin to mechanical aptitude;[2] at least one involving the relation of conjunction, namely, arithmetical ability;[3] and still another in the domain of psychological relations.[4]

It follows, therefore, that an individual who tends to excel in educing relations of one kind will also tend to excel in educing relations of other kinds. But obviously the strength of this tendency will be affected by the presence of such group factors as have been mentioned or others still to be discovered. The field for further research is therefore mapped out. Whenever a group factor is suspected, its existence can be tested by the tetrad technique and, if substantiated, it will be necessary to examine its nature and significance. Thus Ormiston recently obtained evidence for a spatial group factor in the sphere of geography.[5]

FACULTIES

Instead of regarding the field of cognitive operations with respect to the kinds of relations involved, attempts can at least be made to divide it into faculties such as memory or imagination, verbal or musical ability, with a view to ascertain the range and importance of group factors from another point of

[1] C. Spearman, *Abilities of Man*, p. 226.
[2] *Ibid.* p. 229.
[3] *Ibid.* p. 230.
[4] *Ibid.* p. 233.
[5] M. Ormiston, "Psychological Tests and Achievement in Geography", M.Ed. Thesis, Univ. of Leeds, 1933.

view. In this case, uninformed opinion is apt to accept group factors on insufficient evidence. But here again precisely the same technique—analysis by the tetrad criterion—is essential and sufficient. Up to the present the number of group factors for which evidence has been adduced is rather small. Carey found a group factor common to visual and auditory memory;[1] Abelson found a group factor in verbal memories, common to sentences and disconnected words;[2] Krueger and Spearman found one in non-verbal symbolic memories.[3] Carey also found a group factor common to the abilities to form visual and auditory images.[4] At present there is need of further research to ascertain the number of group factors in the field of music.

No discussion of group factors would be complete without reference to the important researches of Kelley. He states that his own findings are on the whole "quite remarkably in harmony" with those of Spearman but adds: "There does seem to be a real disagreement in the importance and extent of a verbal factor and in that of a mental speed factor".[5]

THE SELECTION OF TESTS

A general formula for the correlation of sums or differences due to Spearman[6] shows in a clear light the principles according to which a battery of tests should be assembled.

If there are p tests, namely, a_1, a_2 . . . a_p, and if all the tests have the same standard deviation (cf. Chapter 23), then

$$r_{(a_1+a_2+\,\ldots\,+a_p)g} = \frac{\mathrm{S}(r_{ag})}{\sqrt{p+2\mathrm{S}(r_{aa})}}.$$

The left-hand side of the equation represents the correlation (cf. Chapter 23) between the pool of the p tests and the criterion of "g". The numerator $\mathrm{S}(r_{ag})$ denotes the sum of the correlations between each test and this criterion, while $\mathrm{S}(r_{aa})$ denotes the sum of all correlations between different pairs of the p

[1] C. Spearman, *Abilities of Man*, p. 287.
[2] *Ibid.* p. 288. [3] *Ibid.* [4] *Ibid.* p. 238.
[5] T. L. Kelley, *Crossroads in the Mind of Man*, 1928, p. 23.
[6] C. Spearman, "Correlations of Sums or Differences", *Brit. Journ. of Psychol.* vol. 5 1913, p. 419.

tests. Thus, if $p=3$ and $r_{ga_1}=0{\cdot}4$, $r_{ga_2}=0{\cdot}3$, $r_{ga_3}=0{\cdot}3$ and $r_{12}=0{\cdot}2$, $r_{13}=0{\cdot}15$, $r_{23}=0{\cdot}15$, then

$$r_{(a_1+a_2+a_3)g}=\frac{0{\cdot}4+0{\cdot}3+0{\cdot}3}{\sqrt{3+2(0{\cdot}2+0{\cdot}15+0{\cdot}15)}}=0{\cdot}5.$$

Now as success in some tests may depend more on the general factor "g" than in others, the first principle demands that each test selected should correlate as highly as possible with the criterion of "g" and thus ensure a large numerator.

The second principle may be detected from the fact that $S(r_{au})$ occurs in the denominator, thus it is desirable that the tests should correlate as low as possible with each other. By this is not meant that the correlation would be as low as zero in actual practice as the factor of "g" is common to all the tests, but it is obvious that if any two tests correlated very highly with each other, they would be testing very largely the same thing and it would be unwise to include both.

The third principle demands that a sufficient number of tests[1] (whether 6, 8, 10 or 12) should be chosen so that when the marks assigned to each test after appropriate weighting are added up, the results of applying the tests to a group should agree as reasonably as can be expected with the recognised criterion.[2] It would not be worth while to include a seventh test if the results with seven tests did not show an appreciably higher correlation with the criterion than that obtained with six tests. Complete agreement with the criterion would, of course, not be expected. Thus even if the criterion is the estimate of a highly reliable teacher, it would be coloured by many other factors such as the pupils' personal appearance, his sense

[1] Professor Piaggio has emphasised the importance of the first principle of high "g"-saturation. The above formula shows that increasing the number of tests will have little effect without high "g"-saturation. Cf. H. T. H. Piaggio, "Conditions for a Real and Unique 'g'", *Brit. Journ. of Psychol.* vol. 24, 1933, p. 104.

[2] Appropriate weighting gives a score to an individual x for a whole team of z tests proportional to

$$G_x=w_a\cdot m_{ax}+w_b\cdot m_{bx}+\cdots+w\cdot m_{zx},$$

where the weighting w_u of the mark m_{ux} in any test u is made proportional to $r_{ug}/(1-r^2_{ug})$. C. Spearman, *The Abilities of Man*, Appendix, XIX.

of humour, his politeness, his industry, his emotional and his moral characteristics. And if the criterion were school marks or examination results, these do not depend on "g" alone but also on the teacher's personality, the excellence of his teaching methods as well as on the pupil's interest and industry and attendance at school. Terman concludes that "lack of self-confidence, personal traits which tend to cause over-rating or under-rating, mental inertia, physical defects, emotional instability and psychopathic heredity are the most common causes of discrepancy between mental age and quality of school work".[1]

It is not therefore surprising that a correlation greater than 0·7 is not obtained between intelligence as measured by a mental test and as gauged by a teacher or assessed from school marks. Yet such considerations do not prevent an efficient battery of tests from correlating very highly with the general factor "g".

RELIABILITY OF A TEST

It is advisable to have two equivalent forms of each type of test selected. The correlation between the two forms has been called by Spearman the "reliability coefficient", and the higher it is the better. Thus suppose a group of children is tested with a list of "Opposites" today and with another equally difficult list tomorrow, the two orders of merit would not be expected to be quite the same. There might be *regular* variations due to practice, fatigue, etc., which might be controlled by grouping the marks, *e.g.* the 2nd half of the 2nd test plus the 1st half of the 1st test might be compared with the 2nd half of the 1st plus the 1st half of the 2nd. There would also be *accidental* variations which, unless they were slight, would baffle control and render the test practically useless. If, however, the reliability coefficient is 0·7 or higher, we can at least say that the test measures something with some consistency, although further experimentation would be necessary to prove that it is the general factor which is tested. Obviously a test such as the Strength of Grip Test may have high reliability, say 0·9, and

[1] L. M. Terman, *The Intelligence of School Children*, 1921, p. 106.

yet possess little or no value as a test of intelligence. There is therefore the further need of the concept of validity, but before discussing it, it will be convenient to expound another important concept:

OBJECTIVITY OF A TEST

Although the reliability coefficient may be high when the two forms are applied to a group by a particular tester, it does not follow that the results would even approximate to those obtained by another tester, who might have introduced variations not only in applying the test but also in the marking. In that case the test is not sufficiently *objective* and is at the mercy of subjective differences of the different testers. In practice, then, a test cannot be considered objective unless the results of two testers, who have applied the test to the same group, show high correlation. It follows then that the objectivity of a test is usually lower than its reliability.[1]

VALIDITY OF A TEST

The *validity* or *predictive value* of a test is merely the extent to which it does measure what it purports to measure. This is determined by finding the correlations between the test and some independent criterion. Thus the validity of a new test of intelligence would be determined by correlating it with the estimates of intelligence given by a reliable teacher, or by correlating it with school marks or with the results obtained by means of an intelligence scale of reputable validity.

High correlations afford the best available evidence that tests are measuring the same thing as the criterion. It is obviously important to secure a criterion which is as reliable and objective as possible. As Bingham[2] points out: "The available criterion may be unreliable because records are incomplete, or inaccurate, or difficult to express in comparable units".

It follows that if the validity of a test is high, there is no need to trouble about its reliability, for it must of necessity be

[1] Cf. W. A. McCall, *How to Measure in Education,* chap. 11, 1922.

[2] W. V. Bingham, "Reliability, Validity and Dependability", *Journ. of Appl. Psychol.* vol. 16, 1933.

high. It also follows that the validity of a test cannot be higher than the reliability of the test or of its criterion, and it may be much lower.

DEPENDABILITY OF A TEST

The question now arises: When is the validity of a test sufficiently high so that considerable dependence can be placed on the score of a particular individual as his measure of the trait in question? The correlation between the test and the criterion has to be very close—much closer than usually obtains in practice—in order to satisfy this condition. By means of a coefficient of dependability it will be possible to ascertain how much better than mere chance is prediction from a single score. Kelley's coefficient of alienation is $k = \sqrt{1 - r^2}$, and the dependability coefficient may be regarded as $D = 1 - k$ or $1 - \sqrt{1 - r^2}$. By plotting the curve connecting r and D it will be seen that an r of 0·7 gives only low dependability and that D does not become high until $r = 0{\cdot}9$.

THE VALIDITY OF THE BINET-SIMON SCALE.—The principles enumerated may also be applied to determine the validity of individual tests of intelligence such as the Binet-Simon Scale. There are several methods available.

(1) Burt adopted as his criterion the considered estimate of an observant teacher working daily with the children over a period of several months or years. For each age the correlation between the Teachers' Estimate and the Binet-Simon result was ascertained. In ordinary elementary schools the Binet-Simon tests, as tests of intelligence, prove but moderately successful. For the older children Burt found his own Reasoning Tests to correlate 0·70 with the teachers' estimates whereas the Binet-Simon tests only correlated 0·51 with the teachers' estimates.[1]

(2) Another method is to test the validity of each test in the Binet-Simon Scale by comparing the percentages of success among normal and defective children respectively. The coefficient of colligation may be recommended for this purpose

[1] C. Burt, *Mental and Scholastic Tests*, p. 200.

(cf. Chapter 23). By its use Burt found that many of the individual tests have but comparatively little value, while others are highly diagnostic.[1]

(3) Somewhat similar was the method employed by McCrae,[2] who sorted out the Binet-Simon tests according as they seemed to involve more essentially eduction or reproduction. The following questions were asked in the case of each test: (a) Does the test essentially involve the eduction of relations or correlates or both? (*b*) Are the material and form of the test the best obtainable for this purpose? (*c*) Are the material and form of the test of such a nature that, as near as may be, we are testing these eductions, and them only for the first time? By the use of these questions, it was found possible to make a theoretical division of the tests into four classes: those to be retained, those which might possibly be retained, those which might possibly be rejected, those to be rejected.

The tests were then applied to children of two special sorts: (1) mentally defective who had enjoyed the full ordinary amount of time at school, (2) physically defective whose schooling had been greatly curtailed. Clearly the most satisfactory tests of innate ability should be those in which superior natural endowment enabled the physically defective children furthest to outstrip the mentally defective children despite inferior educational opportunity, while the least satisfactory tests would be those in which superior educational opportunity enabled the mentally defective children to approach nearest to the physically defective. For each age from seven to fourteen years the percentage of M.D. and of P.D. children passing each test was calculated. The percentage difference between these two figures was then calculated. Although the practical testing did not, in all cases, bear out the theoretical finding, there was a very large measure of agreement. Concerning some forty of the Stanford-Binet tests the coefficient of association between the *a priori* judgment and the practical judgment was found to be 0·89. Thus in the case of the test of stating similarities between common objects, *e.g.* "In what way are

[1] C. Burt, *Mental and Scholastic Tests*, p. 204.
[2] C. R. McCrae, *loc. cit.*

wood and coal alike?", it was decided *a priori* to retain it. 28 per cent of M.D. and 60 per cent of P.D. children passed the test. A percentage difference of 114. On the other hand, the test of defining common objects in terms superior to use was on *a priori* judgment rejected, as statements in the form of genus and differentia are essentially a school product and involve but little eduction. 32 per cent of M.D. and 43 per cent of P.D. children passed the test. A percentage difference of only 34.

THE PRACTICAL VALUE OF TESTS OF INTELLIGENCE IN SCHOOLS

One of the greatest needs in education today is to secure a sufficient number of teachers trained to use tests. It is true that the modern group test can often be given with a minimum of training, but no one can hope to use an individual scale such as the Binet-Simon without possessing a considerable psychological background, and it is further very desirable that he should attend a special course on Binet testing.

Individual and group tests have each their advantages as well as disadvantages. Briefly, the individual scale may be used to test the child who may be backward enough to require special treatment; thus a child with an I.Q. below 70 may need to attend a special school. It may also be used with the brilliant child who should be promoted and encouraged to sit for a scholarship. Thus a child with an I.Q. of 125 may possibly, and one with an I.Q. of 135 may probably, win a secondary school scholarship. A child with an I.Q. of 140 may possibly, and one with an I.Q. of 150 may probably, win a university scholarship.

Individual testing is also useful in the case of the emotionally unstable child or the very young child, for in such cases group testing is not so reliable.

Group tests enable large classes to be tested in a short time and are specially adapted for ensuring that children are placed in classes suitable to their mental ages and thus prevent the extreme overlapping of mental ages which Terman[1] and others found in the different grades. They also enable an inspector to

[1] L. M. Terman, *The Intelligence of School Children*, 1921, p. 25.

gauge the general mental ability of different classes and thus to know what progress to expect from each.

Other uses are to study the relation between intelligence and social status,[1] or between intelligence and occupation,[2] or between intelligence and size of family.[3]

It is obvious from what has been said that by means of individual or group tests, or both, considerable educational guidance may be given to a child, especially in a central or secondary school as to the courses of study from which he is likely to profit. Further, vocational guidance is largely based on performance in the tests. As an instance of the use of the Stanford Revision of the Binet-Simon tests may be cited the work of Brier.[4] Candidates for admission to a secondary school were given a test of attainment in Reading, Arithmetic, English, History and Geography occupying six hours. They were mostly between the ages of eleven and twelve, but candidates up to the age of thirteen were allowed to sit. The pass list consisted of 114 children, and each was tested by the Stanford tests. It would appear from the table that the formal examination in school subjects is not in itself a satisfactory method of assessing a pupil's ability to profit by a course at a secondary school. It may admit a dull child and may exclude a child of high intelligence. In the above case the formal examination obviously favoured the over-age candidate.

I.Q.	Frequency.		
	Nos. 1 to 38.	Nos. 39 to 76.	Nos. 77 to 114.
Up to 89 . .	1	1	10
90 to 109 . .	19	24	23
110 to 119 . .	9	11	3
120 and above .	9	2	2

[1] C. Burt, *Mental and Scholastic Tests*, 1921, p. 190; J. F. Duff and G. H. Thomson, *Brit. Journ. of Psychol.* vol. 14, 1923, p. 192.

[2] H. Macdonald, *ibid.* vol. 16, 1925, p. 123.

[3] H. E. G. Sutherland and G. H. Thomson, *ibid.* vol. 17, 1926, p. 81.

[4] A. Brier, "A Critical Inquiry into the Practical Value of Mental Tests in Secondary Schools". M.Ed. Thesis, Univ. of Leeds, 1924.

In some secondary schools pupils on admission are divided into three first-year forms, then at the end of the first year these pupils are reorganised on the basis of the Midsummer Term Examination. Children with an aptitude for science and mathematics go to Form IIb; those who show special ability in languages are placed in Form IIa. Form II consists of the lowest third of the Midsummer Term Examination List. Pupils in IIb pass into IVb. and thence into Vb, the School Certificate Form, whilst those in IIa pass through IVa into Va, the other School Certificate Form. Pupils in II have a more extended course passing into III and IV and then into Vb. As only the "a" Forms take Latin, pupils cannot be transferred from IV to Va. Such a procedure shows the vital necessity of taking all precautions that no pupil should be placed in Form II by mistake. Alternately it may be argued that a more elastic system should be introduced whereby transfers could be effected more easily.

COMPARISON OF INTELLIGENCE SCALES.—At the present time when so many intelligence scales are available, it is necessary to inquire whether and how far they can be regarded as measuring the same thing, or how far would the results obtained by the use of one be expected to agree with those from another.

In 1922 Wilson[1] conducted a careful comparison of several well-known scales, viz. the Terman, Otis, National, Northumberland (Thomson), and Simplex. 340 children were tested. In addition to ascertaining the degree of agreement between any two of these scales, Wilson also found how the results with each scale compare with examination marks, and also with teachers' estimates of intelligence. The discriminatory capacity of each scale for each age and standard was also carefully investigated, and the several tests in each scale were analysed. The results obtained with the different scales were found on the whole to correlate highly with each other, or about 0·7 to 0·8. The scales correlated with examination marks to the extent of 0·39 up to 0·49. With teachers' estimates of intelligence the extent of the correlation was only 0·30. In view of the unreliability of the

[1] J. H. Wilson, "A Survey, Historical and Critical, of Intelligence Tests", M.Ed. Thesis, Univ. of Leeds. 1923. Cf. also his "Comparison of Certain Intelligence Scales", *Brit. Journ. of Psychol.* vol. 15, 1924.

estimates as compared with tests such a result is not surprising.

Intelligence and Quickness.—Intelligence tests have been criticised on the score that the time-limits have been too severe and that they may measure a kind of quickness but not intelligence. The writer once asked a famous admiral if he did not think it possible for a person to be very intelligent but slow. His answer was characteristic: "We do not want him in the Navy". When the question was put to Professor Spearman another characteristic answer was obtained: "Test it and see". Even if popular opinion would answer with a decided affirmative, it was nevertheless a question for careful investigation. To complicate matters further it is alleged that there are sparkling, quick-witted individuals who fail to distinguish themselves in intelligence tests, also individuals who are slow starters in any task, and still others who suffer from nervousness. The question has a serious aspect. If pupils sit an intelligence test for secondary school scholarships, and if the marks are added to the English and Arithmetic marks, there is still a grave risk, according to a recent enquiry by Professor Valentine,[1] that a pupil who deserves a scholarship may not obtain it. But the available evidence gives little support to the view that this injustice is owing to some being very intelligent but slow. It is well to note that some major causes have been suggested by Valentine, but are not here our concern.

Miss M. D. Kay[2] tested twenty-four "bright" girls and twenty-four "dull" girls, using well-known types of group intelligence tests. The time allowed for each test was very liberal. Equivalent tests were afterwards given where the time allowance was too severe to enable more than one pupil to finish any test. Very high correlations were obtained between the two series of tests both in the case of the "bright" and also in the case of the "dull". Such a result offers no support to the conclusion that there are intelligent pupils who are slow. Bernstein[3] made a thorough study of the relation between

[1] C. W. Valentine, *The Reliability of Examinations*, 1932.

[2] M. D. Kay, B.A. Thesis, Univ. of Leeds, 1923.

[3] E. Bernstein, "Quickness and Intelligence", *Brit. Journ. of Psychol.* Mon. Supplt. No. 7, 1924.

quickness and intelligence and came to the conclusion that there was no speed ability, independent of general ability. In other words, there was no evidence of a subject being slow and yet intelligent, or quick and yet dull.

INTELLIGENCE AND SCHOLASTIC ACHIEVEMENT OF THE CRIPPLED CHILD

164 crippled children from the age of nine years upward were the subjects of a recent research by Garry.[1] Three groups specially studied were (*a*) heart disease, congenital rheumatism, valvular disease; (*b*) infantile paralysis, spastic paralysis; and (*c*) non-pulmonary tuberculosis. Their I.Q., Arithmetic Quotient and English Quotient were obtained by means of Burt's Northumberland Tests. All the testing was done between June and October 1929. No significant correlation was shown between I.Q. and attendance. This is contrary to the conclusion reached by Gordon,[2] but as Garry's subjects were older, they may have obtained sufficient mastery over reading and arithmetic to enable them to make progress when absent from school. Their average for arithmetic was lower than that for English, and this is not surprising for owing to their inactivity they may devote comparatively more time to reading. On the whole, children suffering from cardiac rheumatism scored more highly than the others. A suggested explanation is that this disease generally attacks children over seven years. Children suffering from the various forms of paralysis had the lowest I.Q., and often their inability to control the muscles used in speech and writing handicaps their powers of expression so that they are judged less intelligent than they really are.

The average I.Q. of all the children tested, $88{\cdot}6 \pm 0{\cdot}84$, is roughly the same as that obtained by Gordon, or Dawson,[3] who tested physically defective children by means of individual tests. In Dawson's investigation the diseases were not mainly

[1] N. Garry, "The Education of the Crippled Child in England", M.A. Thesis, Univ. of Leeds, 1933.

[2] H. Gordon, *Mental and Scholastic Tests among Retarded Children*, Bd. of Educ. Pamphlets, No. 44.

[3] S. Dawson, assisted by J. C. M. Conn, *Intelligence and Disease*, Med. Res. Council. Publ., 1931.

crippling diseases, and on the whole it was only disease of the ductless glands and of the brain which affected the intelligence.

VERBAL AND NON-VERBAL TESTS OF INTELLIGENCE

Sometimes non-verbal tests are used as a matter of necessity, *e.g.* to test immigrants who have little command over the language used in the test. But, apart from this, it is necessary to investigate (1) whether the ability measured by the usual verbal tests of intelligence is the same as that measured by certain non-verbal tests, and (2) if it is the same, which gives the truer measure.

Important researches by Davey,[1] Line,[2] Fortes,[3] and Stephenson,[4] from the laboratory of Professor Spearman have thrown much light on these matters. Davey concluded that a verbal mental test measures the same general factor "g" as does a test similar in form but non-verbal in material, *e.g.* pictures. Later, Line constructed a series of non-verbal perceptual tests. Confining himself to the eduction of relations he discovered that his perceptual tests correlated as much as 0·89 with "g". Here, then, is the possibility of constructing a truly international test of intelligence which promises to be an improvement on the usual types of performance scales of intelligence. Mention must also be made that considerable progress in this direction had also been made by Dodd[5] in America. Fortes carried the work of Line a step further by employing tests which were more complex in form and demanded the eduction of relations and correlates. His results corroborated those obtained by Davey and Line using other perceptual tests. Finally, some recent experiments by Stephenson corroborated those of Davey, Line and Fortes as to the absence of a

[1] Constance M. Davey, "A Comparison of Group Verbal and Pictorial Tests of Intelligence", *Brit. Journ. of Psychol.* vol. 17, 1926.

[2] W. Line, "The Growth of Visual Perception in Children", *Brit. Journ. of Psychol.* Mon. Supplt. vol. 15, 1931.

[3] M. Fortes, "A new Application of the Theory of Noegenesis to the Problem of Mental Testing", Ph.D. Thesis, Univ. of London, 1930. Cf. also "Perceptual Tests of General Intelligence for Inter-Racial Use". *Trans. of the Roy. Soc. of South Africa*, vol. 20, Part 3, 1932.

[4] W. Stephenson, *Journ. of Educ. Psychol.* vol. 22, pp. 167-185, 255-267, and 334-350, 1931.

[5] S. C. Dodd, *International Group Mental Tests*, Princeton, N.J., 1926.

factor common to the non-verbal sub-tests other than "g". But, on the other hand, he found indications of a group factor extending throughout the verbal sub-tests.

Gowda[1] employed a battery of eight non-verbal tests of intelligence, namely, Identity, Digit-Symbol, Most Unlike, Analogies, XO Series, Thurstone A, Thurstone B and Fouracre's Geometry Test which proved suitable for children aged eleven and twelve, the correlation with "g" being 0·90. On working out the 210 tetrad differences it was found that the tetrad criterion was satisfied.

Gowda also employed verbal tests of intelligence, namely, Tomlinson's "West Riding" Scale as well as Karve's Fluency tests with the same subjects and the evidence suggested the presence of a factor, other than "g", common to the verbal tests of intelligence and the fluency tests, and this factor appeared to be more on the associative than on the imaginative side of fluency.

The further conclusion was drawn that the factor or factors involved in fluency tests and verbal tests of intelligence make small but significant contributions to success in school, examination marks in English, Arithmetic, History, Goegraphy and Science being pooled as a measure of school achievement. There is need of further research into the nature of both the "fluency" and the "verbal" factor.

This need is emphasised when it is pointed out that pupils from poor homes or districts, when tested by the usual verbal type of test, may occasionally be handicapped owing to their relative backwardness in verbal ability. This danger is at least reduced by the use of non-verbal tests. But it is not to be concluded that the distribution of intelligence in the community, when measured by non-verbal tests, will be found very different from that given by means of verbal tests.

Sumithra[2] employed a battery of non-verbal tests of intelligence in "superior", "average" and "inferior" districts. In the schools selected the distribution was similar to that found

[1] A. C. D. Gowda, "Some Mental Determinants of Scholastic Achievement", M.Ed. Thesis, Univ. of Leeds, 1932.

[2] D. Sumithra, "An Experimental Investigation into the Distribution of Intelligence in Different Schools by Means of Non-verbal Tests", M.Ed. Thesis, Univ. of Leeds, 1933.

by investigators who used verbal tests. Thus Tomlinson,[1] using verbal tests, found the percentage of boys who had an I.Q. of 100 or above to be 68, 48 and 19 in "good", "average" and "poor" schools respectively.

Sumithra's Spatial Overlapping Test is an example of a test which may be regarded as non-verbal once the instructions have been understood. It is a modification of the simpler test of Stephenson [2] or indeed of the well-known test of Abelson,[3] another pupil of Spearman. "The overlapping figures consist of some or all of these—triangle, square, rectangle, circle and pentagon (five-sided figure). Below the pair all the five figures are given in small sizes. You are to draw a line across these small figures *which do not contain the cross in either of the pair above.*"

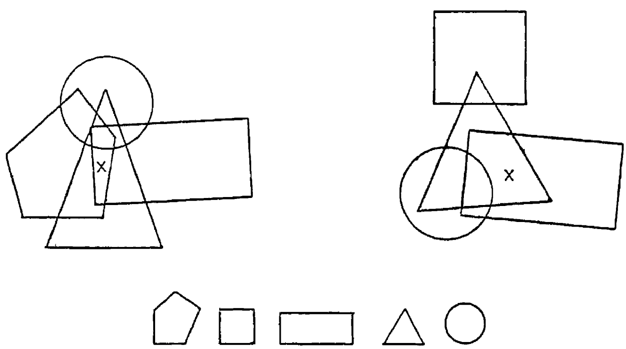

FIG. 17.

Another non-verbal test, not yet published, was devised by Islam. It is called the Moving Piece Test.[4]

QUESTIONS

1. Below is Eifion Wyn's "Englyn" to Heather Flowers. Assuming that you are not acquainted with the character-

[1] T. P. Tomlinson, *op. cit.*
[2] W. Stephenson, *op. cit.*
[3] A. R. Abelson, "The Measurement of Mental Ability of Backward Children", *Brit. Journ. of Psychol.* vol. 4, 1911.
[4] F. Islam, "An Experimental Investigation into the Comparative Validity of some Non-Verbal Tests of Intelligence", M.A. Thesis, Univ. of Leeds, 1934.

istics of an "englyn" nor with the Welsh language, write down all the peculiarities which you note in its structure:

> "Tlws eu tw, liaws tawel,—gemau teg
> Gwmwd haul ac awel;
> Crog glychau'r creigle uchel,—
> Fflur y main,—phiolau'r mel."

What relations were educed?

2. Do you know the exact meaning of every word in the passage given below? If not, state what you do not know or are in doubt about.

> "A tiger with tastes anthropophagous
> Felt a yearning inside his oesophagus,
> He spied a fat Brahmin
> And cried 'Where's the harm in
> A peripatetic sarcophagus.' "

(This method of self-criticism is an excellent way of mastering a psychological textbook, especially when several join in; not only is the extent of ignorance ascertained but also the extent of self-awareness of ignorance.)

3. Examine the contention that the Binet-Simon Scale is too verbal to constitute a reliable measure of general intelligence.
4. To what extent does ability in educing one kind of relation correspond to the ability in educing another kind?
5. A boy, aged 6 years 10 months, said: "The oval relation to a square is an oblong". (By "relation" he meant a relative, *e.g.* a brother.) Consider how the boy may have educed the statement.
6. Examine the following queries of a boy, whose age is given in brackets, with reference to the possible kinds of relations which he may have educed.
   (*a*) "How did the parsnips get dirty?" When told they were obtained from the ground, he asked: "Why do they not grow on trees?" (3 yrs. 6 mo.)
   (*b*) In his bath he asked: "Why is the soap brown and the soap coming off it white?" (3 yrs. 6 mo.)
   (*c*) "Why is there wax in your ear?" On being told, "To prevent dirt from going to the part you hear with", he

then asked: "Is it the same part as you hear with the other ear, or is there another for the other ear?" (3 yrs. 7 mo.)

(*d*) "Why haven't watches got alarums?" (4 yrs. 1 mo.)

(*e*) "If 8 comes after 7, why is it that Mrs. X's house (*i.e.* No. 9) comes after ours (*i.e.* No. 7)?" (4 yrs. 2 mo.)

7. Consider how a boy, aged 4 years 1 month, comes to use the phrase "very very sometimes" to mean "rarely".
8. Discuss the following statement: The measure of a child's intelligence is his ability to educe relations and correlates.

Note.—Examples of children's "Why" questions may be studied in (1) J. Sully, *Studies of Childhood,* 1909 ; (2) W. Stern, *Psychology of Early Childhood,* 1924; (3) J. Piaget, *The Language and Thought of the Child,* 1926; (4) J. Piaget, *Judgment and Reasoning in the Child,* 1928; (5) J. Piaget, *The Child's Conception of the World,* 1929; (6) J. Piaget, *The Child's Conception of Causality,* 1930; (7) J. Piaget, "Retrospective and Prospective Analysis in Child Psychology," *Brit. Journ. of Educ. Psychol.*, vol. 1, 1931; (8) "The Scientific Interests of a Boy in Pre-School Years by Two Parents", *Forum of Education,* vol. 6, 1928; (9) K. Bühler, *The Mental Development of the Child,* 1930; (10) S. Isaacs, *Intellectual Growth in Young Children,* with Appendix on Children's "Why" questions by N. Isaacs, 1930.

References

1. C. Spearman, "'General Intelligence' Objectively Determined and Measured", *Amer. Journ. of Psychol.*, vol. 15, 1904.
2. C. Burt, "Experimental Tests of General Intelligence", *Brit. Journ. of Psychol.*, vol. 3, 1909.
3. C. Burt, "Experimental Tests of Higher Mental Processes and their relations to General Intelligence", *Journ. of Exp. Ped.*, vol. 1, 1911.
4. B. Hart and C. Spearman, "General Ability, its Existence and Nature", *Brit. Journ. of Psychol.*, vol. 5, 1912.
5. C. Burt, *Mental and Scholastic Tests,* 1921 (3rd impression, 1927).
6. C. Spearman, *The Nature of "Intelligence" and the Principles of Cognition,* 1923.
7. C. Burt, *The Mental Differences between Individuals,* 1923.
8. C. Fox, *Educational Psychology* (chap. on Mental Tests), 1925.
9. C. Burt, *The Measurement of Mental Capacities,* 1927.

10. C. Spearman, *The Abilities of Man*, 1927.
11. B. S. Burks, "The Relative Influence of Nature and Nurture upon Mental Development". *27th Year Book of the National Soc. for the Study of Education*, 1928.
12. F. N. Freeman, K. J. Holzinger and B. C. Mitchell, "The Influence of Environment on the Intelligence, School Achievement and Conduct of Foster Children", *ibid.* 1928.
13. *Psychologies of 1930*, edit. by C. Murchison (chap. by C. Spearman, On "G" and after—A School to end Schools), 1930.
14. C. Spearman, *Creative Mind*, 1930.
15. E. M. Lawrence, "An Investigation into the Relation between Intelligence and Inheritance", *Brit. Journ. of Psychol.* Mon. Supp. vol. 16, 1931.
16. H. T. H. Piaggio, "The General Factor in Spearman's Theory of Intelligence", *Nature*, January 10, 1931.
17. R. Pintner, *Intelligence Testing: Methods and Results*, 2nd ed., 1931.
18. G. H. Thomson, *Instinct, Intelligence and Character*, 2nd ed., 1932.
19. Ll. Wynn Jones, "Present-day Theories of Intellectual Factors (General, Group, and Specific)", *Brit. Journ. of Educ. Psychol.*, vol. 3, 1933.
20. W. Brown and W. Stephenson, "A Test of the Theory of Two Factors", *Brit. Journ. of Psychol.*, vol. 23, 1933.
21. *The Intelligence of Scottish Children*, 1933.

CHAPTER 9

A SCALE OF WIT

MATERIALS.—Fifty typed copies of Series A and of Series B, each consisting of eighteen stories. Suitable subjects—Standard VII. children.

INSTRUCTIONS (to be given orally by E to the class).—"There are 18 little stories in the test I am going to give to you. Most of them are jokes taken from newspapers or humorous magazines. Your task is to write down briefly on your answer sheets what is the point of each story. In other words, what exactly in each story makes it a joke. Some can see the point of a joke much better than others. You may think that some of the stories are poor jokes, but even so, try to write down exactly why each is supposed to be amusing. If you do not see any point in a story, do not spend too much time on it, but pass on to the next. Remember to number your answers correctly. Do not waste time trying to write neatly. Make your answers short, even one sentence is often ample. *Do not write anything on the question paper*. If you have heard any of the stories before, put a big cross to the left of your answer. I shall now write a story on the blackboard to make the test clear:

> The rabbit multiplies very rapidly but it takes a snake to be an adder.

Put your hands up all those who see the point of this story." (E should wait until the majority have put up their hands before asking a particular pupil to answer.)

"I shall now write another story in the same way:

> An old beggar had a board suspended from his neck, inscribed: 'Blind from my birth'. Another beggar, reading the inscription as he passed, was heard to remark: 'Well, well! There's a chap who started young in business.'"

(E again gets a show of hands as before.)

Answer sheets are then distributed and each pupil writes his name, standard and exact date of birth at the top.

Series A is then distributed face downwards. At the signals "Ready. Go!", the pupils turn over the question papers and start. The time allowed is thirty minutes. After the signal "Stop" is given the papers are collected.

Series B should then be done in the same way, but illustrative examples are now unnecessary. The answers should then be marked by the students, but it is advisable to agree beforehand on a system of marking. Two marks may be awarded for each satisfactory answer and one mark only for a mediocre. Cases sometimes occur when it is difficult to say whether the point of the story has been apprehended, but on the whole the tests were fairly easy to mark even if they could not be marked with the rapidity which is associated with group tests of intelligence. But the student is amply repaid for the extra expenditure of time by the greater insight gained into the children's outlook. Greater still is this insight when individual children are asked to read these stories. It is then also possible to ascertain when they are handicapped through ignorance of some word or phrase occurring in the context.

Note.—When the tests are taken with adult subjects, the following alternative method is suggested. Let them be divided into pairs of Es and Ss. E directs S to read each story in Series A, only one story to be visible at a time, and to put up his hand immediately he sees the point of the story. The time, taken from the point at which he finishes the reading until he puts up his hand is measured with a stop-watch. For series B, E and S interchange. It is, of course, advisable for E to ascertain what S regards as the point of each story.

SERIES A

1. A dyer, in a court of justice, being ordered to hold up his hand that was all black; "Take off your glove, friend", said the judge to him. "Put on your spectacles, my lord", answered the dyer.
2. A lady's maid told her mistress that she once swallowed several pins together. "Dear me!" said the lady, "didn't they kill you?"

3. When Albert Smith—a well-known lecturer—signed a newspaper article with the initials A. S., Douglas Jerrold asked him why he had told only two-thirds of the truth.
4. At a shop-window in the Strand there appeared the following notice: "Wanted, two apprentices, who will be treated as one of the family".
5. Lady Customer: "How much are grouse today?" Poulterer: "Twelve shillings a brace, ma'am. Shall I send them?" Lady Customer: "No, you need not send them. My husband's out grouse-shooting and he'll call for them as he comes home".
6. The mother of a student sent her son a new black overcoat with a letter in the pocket, which began thus: "Dear John, look in the pocket of the coat and you will find this letter".
7. A lady said to Dr. Johnson that a piece of music which was being played was very difficult. He replied: "I wish it had been impossible".
8. After reading Rousseau's Ode to Posterity, Voltaire remarked: "This poem will not reach its destination".
9. When O'Connell was walking by the River Wey with a friend, the latter remarked that a plant he had just taken from the river was a very rare one. O'Connell said: "It is an out of the way one at any rate".
10. A Scotchman went to see Niagara Falls. "Did you ever see anything so grand?" demanded his friend. The Scotchman replied: "Well, as for grand, I maybe never saw anything better; but for queer, man, do you know, I once saw a peacock with a wooden leg".
11. Dean (who is about to be made a bishop, but who always travels 3rd class): "I once was a curate, my friend".
    Farmer: "Drink, I suppose".
12. At a dinner-party a Mr. Breach requested Tom Hood to pronounce his name as a disyllable, thus, Bre-ak. A few minutes later Hood said: "Mr. Bre-ak, would you be so good as to re-ak me a pe-ak?"
13. "Madam," said the dignified gentleman, "your dog bit me on the ankle." "He did", cried the lady. "Oh, I must send for a doctor!" "Oh, I assure you it isn't as bad as ——" "You're the third person he's bitten today", broke in the lady; "I'm sure he can't be feeling well."
14. A Norfolk man has been fined for beating a boy with a codfish. He shouldn't have used a codfish.
15. Notice in a baker's window: "I am unable to reduce the cost of my bread without reducing its quality. This I cannot do."

16. In the parlour of a public-house in Fleet Street there used to be written over the chimney-piece the following notice: "Gentlemen learning to spell are requested to use yesterday's paper".
17. A commercial traveller came down to breakfast in a country hotel, and the waitress said: "There is beans, pork, cranberry pie, mincemeat, and apple pie". The commercial traveller said he'd take beans, pork, cranberry pie and mincemeat. "Say, mister," said the waitress, "tell me what is wrong with the apple pie."
18. The following notice is to be found in an electric station in Bristol: "Beware! to touch these wires is instant death. Anyone found doing so will be prosecuted."

SERIES B

1. Patrol-leader to his Scouts: "When I say 'one' rise on your toes; when I say 'two', bend your knees; when I say 'three', jump in the air; when I say 'four', you come down again".
2. A newsboy inquired at a nursing-home: "Do you take in the *Daily Mail*?" To which query the new maid replied: "No, we take in the weakly female".
3. A governor of a city bought a tomb, and said, "No living soul shall be buried under it, save those of my own family".
4. A Philadelphia Banker distinguished himself by giving a supper-party at which monkeys mixed with the guests. To avoid confusion, the guests wore evening dress.
5. Lord Gough, after his victory over the Sikhs, was asked what were the relative strengths of the opposing armies. "They were Sikhs, and we won", replied he.
6. An English admiral once ate an orange after breakfast, instead of before it, as is the American custom. One American lieutenant said: "Admiral, why don't you eat your orange before breakfast?" Another American replied: "Oh, that's what William the Conqueror used to do".
7. "Sirrah," said a judge, to one brought before him, "you are an arrant knave." Said the prisoner: "Just as your Worship spoke, the clock struck two".
8. Father (reprovingly): "Do you know what happens to liars when they die?" Johnny: "Yes, sir; they lie still".
9. An Englishman entered an American Restaurant with the temperature 60 below zero and said, "Waiter, bring me an iced lemonade". The waiter replied, "We're out of lemonade; how would you like a pair of linen trousers?"

10. FATHER: "Goodness, what have you done, child?" FRANK: "I fell into the canal". FATHER: "What! With your new trousers on?" FRANK: "I didn't have time to take them off".
11. It is said of the dwellers on the banks of some of the rivers of British Columbia, concerning the fish they catch, that what they can they eat, and what they can't they can.
12. GENERAL (at manœuvres): "Why didn't you send your men to the left?" LIEUTENANT: "I gave the order, Sir, but apparently I've been misunderstood". GENERAL: "Sir, a man who can't make his subordinates understand him is a fool. Do you understand me?" LIEUTENANT: "No, Sir".
13. The following epitaph was once seen in Ireland: "Erected to the memory of John Moran, accidentally shot as a mark of affection by his brother".
14. Somebody asked why B stood before C. "Because", said his friend, "a man must B before he can C."
15. "Have you the rent ready?" "No, sir; mother's gone out washing and forgot to put it out for you." "Did she tell you she'd forgotten?" "Yes, sir."
16. A passenger waiting for a train at a station, noticed a cat without a tail. "Manx?" he inquired, pointing a finger at the animal. "No", said the stationmaster, "twelve-thirty express."
17. The horse ridden by an Irishman began kicking violently and at last caught its hind leg in a stirrup. "Faith", said the Irishman, "if you want to get on, it's time for me to get off."
18. A man had just built his own house. "Why, the chimney isn't straight," said his friend, "it leans over to the left". "Yes," said the man, "but if you go round to the back of the house, it leans over to the right, so it must be straight."

DISCUSSION

Wit and humour are notably difficult to define. Mark Lemon's conviction was that the attempt to define wit was like that to define beauty. Bernard Shaw wrote: "There is no more dangerous literary symptom than a temptation to write about wit and humour. It indicates the total loss of both." This indeed if taken seriously would implicate at a stroke most of the greatest figures in philosophy and literature! Thus John Locke distinguished sharply between wit and judgment: "For wit lies most in the assemblage of ideas and putting these together with quickness and vivacity, whenever can be found

any resemblance or congruity, whereby to make up pleasant pictures and agreeable visions of fancy; judgment on the contrary, lies quite on the other side, in separating carefully, one from another, ideas wherein can be found the least difference, thereby to avoid being misled by similitude, and by affinity to take one thing for another."[1]

Hazlitt, however, disagrees with Locke and very plausibly holds that "The shrewd separation or disentangling of ideas that seem the same, or when the secret contradiction is not sufficiently suspected and is of a ludicrous or whimsical nature, is wit just the same as the bringing together those that appear at first sight totally different".[2] An interesting question suggests itself: Does the wit and the appreciator of wit in others tend to be the same person? *A priori*, one might expect the wit to appreciate wit in others, but that the converse need not hold. It is a question on which some light may be shed by experimental methods.

Turning to modern authors there are two in particular who have propounded original views on the nature of wit. Bergson compares the wit with the poet. In a broad sense he maintains that wit is a certain dramatic way of thinking. In every wit there is something of a poet, but whilst poetic creation demands some degree of self-forgetfulness, the wit is not usually capable of this for he is not wholly engrossed in the business because he only brings his intelligence into play. Bergson[3] holds that the poet would turn into a wit by simply resolving to be no longer a poet in feeling, but only in intelligence. His readers are, however, left in grave doubt as to whether the poet could successfully carry out this resolve. In discussing wit in the restricted sense, Bergson contents himself by merely defining wit as a gift for dashing off comic scenes in a few strokes.

Freud subdivides wit into "harmless" and "tendency" wit, the latter including obscene, hostile, cynical and sceptical wit. He claims that wit can be solved by the same mechanisms as

[1] J. Locke, *An Essay concerning Human Understanding*, vol. 1, p. 203.
[2] W. Hazlitt, *Lecture on Wit and Humour*, 1818, p. 19.
[3] H. Bergson, *Laughter*, 1921, p. 106.

dreams and the neuroses, and holds that the pleasure created by wit corresponds to the economy of psychic expenditure in inhibition. Greig, however, points out that "the misconception of wit as economising psychic expenditure arises through confusion of speed with force. We have it on good authority that brevity is the soul of wit. A really effective witticism does its business in a flash. But speeding up behaviour in this way is not the same as economising effort; quite the contrary. More has to be got through in a given time, and that can only be done in human endeavour as with a piece of machinery by increasing the pressure on it. It is precisely for this reason that wit is notoriously fatiguing". [1] Greig, however, agrees with Freud that wit is a compressing tendency. The last two sentences in Freud's remarkable book are as follows: "The euphoria which we are thus striving to obtain is nothing but the state of a bygone time in which we were wont to defray our psychic work with slight expenditure. It is the state of our childhood in which we did not know the comic, were incapable of wit, and did not need humour to make us happy." [2] Such a statement would possibly appear more convincing if elucidated more fully!

For the purposes of the present chapter it was considered unnecessary to accept any exact definition of wit. A collection has merely been made of stories which would be considered to involve sound-wit, play on words, the wit of caricature, characterisation wit, distortion, witty repartee, absurdities and the like. Even the pun was included in spite of the severe strictures of Addison and Johnson. Some may be considered childish almost to the point of inanity, others perplex the majority of university graduates. But it appeared necessary to choose such stories as would cause the child at least a momentary hesitation before the point was seen. For this to the writer seems a characteristic of wit no matter what theory be held concerning it. In addition, the stories should be as short as possible, for not only is brevity the soul of wit but the exigencies of experimentation

[1] J. Y. T. Greig, *The Psychology of Laughter and Comedy*, 1923, p. 216.
[2] S. Freud, *Wit and its Relation to the Unconscious*, trans. by A. A. Brill, 1916, p. 384.

also demand it. Finally, in order to obtain a more genetic point of view, the tests were given to classes differing appreciably in age. Not only were they given to a class of Standard VII. boys in an elementary school but also to a Standard V.(b) boys in the same school as well as to a Form 5A of secondary

Series	31 Graduates	31 Boys, Form 5A	45 Boys, Standard VII.	42 Boys Standard V.(b)
A4	97	100	71	30
B12	97	95	63	13
A6	100	95	83	64
B15	97	94	84	37
A18	94	92	73	67
B2	100	90	42	13
A1	87	87	88	44
B14	100	87	48	13
B10	92	85	74	57
B1	100	84	90	55
B3	97	77	77	46
B11	98	77	28	0
A7	95	77	63	21
B4	97	71	37	4
A2	92	70	67	46
B17	82	66	48	14
A5	89	66	57	6
A13	94	66	43	4
B8	97	63	33	19
A3	78	58	22	1
B5	51	52	14	0
B7	65	48	28	8
B16	47	48	21	0
A9	84	48	21	2
B18	84	45	52	17
A8	94	45	8	2
B9	87	39	27	7
A12	97	37	16	6
B13	97	35	14	5
A15	84	35	20	4
A11	51	34	7	0
A17	74	16	14	10
A14	55	15	7	10
A16	68	10	4	0
B6	28	5	8	0
A10	60	5	8	2

school boys.[1] They were also given to men and women graduates, as even in their case it was rightly presumed that some of the stories would cause some perplexity. The following table shows the degree of difficulty experienced by the various groups in answering each item of Series A and B, the figures denoting the percentage of correct answers.

The following table shows the range of marks in each group:

	Lowest Mark	Lower Quartile[2]	Median	Upper Quartile	Highest Mark
Standard V.(b)	2	7	12·5	15·5	33
Standard VII..	6	24	30	36	54
Form 5A .	25	37	42	50	60
Graduates .	43	54	60	63	68

In Standard V.(b) lack of understanding of the language makes it impossible for several of the pupils to appreciate some of the stories. Thus poor comprehension is responsible for the following answers:

B4: "Foolish because a guast is a spirit" (evidently "guest" was read as "ghost").

B7: "you cannot worship spoke".

Doubtless many in this class who saw the point of some of the stories were yet unable to give it expression in words.

The power of expression, however, improves much by the time Standard VII. is reached. Nevertheless some lack of comprehension is still responsible for many curious answers in the case of the Standard VII. boys and of the secondary school boys, and occasionally even of graduates. The following specimens illustrate how "far-fetched" were some of the answers obtained:

Standard V.(b)

A1: It is funny because the man who was in court put a glove on his black hand so that the dirt would not be seen.

B4: Foolish because a guast is a spirit.

B7: You cannot worship spoke.

B16: Foolish because he can't see the time on a cat's tail.

[1] The writer must express his thanks to Mr. Whatmoor, then Headmaster of the Blenheim School, Leeds, and to Mr. Darling, then Headmaster of the West Leeds High School, for permission to give the tests.

[2] Cf. Chapter 23.

Standard VII.

A5: Lady thought he meant braces you have on your trousers.
A13: She thought that the flesh the dog had taken out of the man's leg was bad.
A13: The dog can be well and bite people at the same time.
B11: You can't eat fish if you have not got it.

Form 5A

A3: Smith when an S has been taken off becomes mith which means a fairy tale.
A3: Douglas Jerrold should not know that it was Albert Smith from only seeing the initials.
A8: Will not go to the poor (presumably confusing poverty and posterity).
A8: Poems cannot walk.
A8: The poem would not have descendants.
A8: He possibly thought it was Road to Posterity.
A10: The peacock might only have found the wooden leg.
A10: The peacock was not born with a wooden leg.
A11: The farmer thought he cured himself of drink.
A14: The boy had the codfish.
A14: "Beating" means "getting it before him".
A14: Cod is rather expensive.
A17: She said apple-pie last and the commercial traveller showed his worst things last.
B5: They were all battling Siki's (great boxer).
B6: He meant William of Orange.
B7: He was not speaking the truth as Peter in the Bible when the cock crowed.
B7: Why wasn't the clock brought up at court, it struck two?
B11: What they can catch they can eat, they can eat anything.
B13: His brother put him out of misery.

Graduates

A8: Posterity has no end or destination, play on word posterity.
A11: Thought curate meant man who cures.
A14: Cruelty to codfish but not to boy.
B4: To avoid confusion between guests and family, though it reads as if between monkeys and guests.
B5: Inferring that the Sikhs were naturally weak.
B5: Sikhs pronounced sick.
B5: This implies that the Sikhs were strong, but as they won they must be stronger.

B6: William the Conqueror died.
B6: Oranges were not discovered then.
B7: No point in prisoner's answer—absurd.
B7: Hadn't heard the judge.
B16: Thought he meant place train was going to.
B16: Absent-minded stationmaster speaks of train.
B16: Stationmaster's mind on train, answer to question he expected given to another.

The conclusion seems to be that when the point of a story is not apparent, the mind in actively searching for a clue often fastens on quite an irrelevant detail which is accepted as the solution on very inadequate grounds. The extent to which this is done might well furnish a measure of S's suggestibility in a very real sense of the word, but it would be hard to disentangle the factor of suggestibility from that of "intelligence".

Another conclusion is that the appreciation of puns or, more generally, of play on words is but little developed in Standard V.(*b*). It leaves much to be desired in Standard VII., and even in Form 5A of the secondary school it is distinctly below the standard attained by most graduates.

Lack of comprehension of words and phrases makes it difficult for most children in Standard V. to appreciate the majority of witty stories of a character similar to those given in these tests.

Before such stories could be suitably chosen for children in this class it is necessary to possess a considerable knowledge of their mentality. These conclusions refer mainly to the written word and must not be taken to imply that children aged four or five cannot appreciate simple play on words in conversation.

[There is still need for investigations to ascertain what meanings are given by children and adolescents to words in common currency. An important investigation by Presswood,[1] the results of which are not yet published, suggests that many adolescent boys, who do not proceed to the secondary school, show little or no advance in their comprehension of words during the two years after leaving the elementary school. If

[1] R. E. Presswood, "An Experimental Study of Concepts in Adolescence." M.Ed. Thesis, Univ. of Leeds, 1934.

the concept-forming process is not sufficiently developed at fourteen years, then the argument for raising the school-leaving age to fifteen receives further justification.]

It is sometimes held that wit is relative, not only to class, but also to locality. Thus Chesterfield wrote that wit is relative to class and locality and that wit, humour and jokes thrive in a particular soil and will not often bear transplanting. It would be possible to check such statements by giving a series of tests to children from widely different localities. A preliminary survey by the present writer seemed to indicate that locality as such plays an insignificant part, if any, in comparison with the paramount factor of "general intelligence".

Another writer[1] goes so far as to state: "A man is born to see a particular sort of joke; or he is not. You cannot educate him into seeing it." Such a statement, however, appears to be belied by the fact that the capacity to enjoy the comic press of some countries can be acquired by foreigners who at first fail to appreciate it.

QUESTIONS

1. Attempt a classification of witty stories.
2. From actual examples of each type of witty story, consider which of Spearman's relations have to be educed, and refer to the correlate finding you find involved in each.

[1] J. Palmer, *Comedy*, p. 5.

CHAPTER 10

FLUENCY OF ASSOCIATION AND IMAGINATION

DIRECTIONS.—"I want you to write as many nouns as you can which begin with a certain letter of the alphabet. You will be allowed only two minutes. I shall pronounce the letter and simultaneously write it on the blackboard, then you should start immediately."

Note.—This test should be repeated with four different letters. Those recommended for use are those which are given in the account of Karve's tests in this chapter.

Each S should write his introspections and refer to the difficulty of the task. As arranged above, the test can be used as a group test, but if taken as an individual test the nouns could be spoken instead of written and the time allowed might be halved.

If time allows, other tests taken from the series by Karve should be performed.

DISCUSSION

The word "fluency" seldom appears in the index of a psychological textbook, yet there are indications of the need for some such word. According to the dictionary "fluency" means a smooth easy flow, especially in speech. And Steele in the *Tatler* writes that "a man of weak capacity, with fluency of speech, triumphs over you". It is of course true that individuals are most fluent when conversing on matters within their interests and that many are struck dumb once the discourse touches on unfamiliar topics, *e.g.* "Everything he speaks smells of gunpowder; if you take away his artillery from him, he has not a word to say for himself".[1] That, after all, is not only natural but sometimes desirable, for who has not met the bore who is equally fluent on all subjects however colossal his ignorance. Steele's and Addison's ridicule of the mere courtier,

[1] Richard Steele and Joseph Addison, *Sir Roger de Coverley*.

the mere soldier, the mere scholar or the mere anything, as insipid, pedantic characters does not mean that the chatterbox would meet with kinder treatment at their hands. Dean Swift[1] also was of opinion that "The common fluency of speech in many men, and most women, is owing to a scarcity of matter, and a scarcity of words; for whoever is a master of language, and hath a mind full of ideas, will be apt, in speaking, to hesitate upon the choice of both; whereas common speakers have only one set of ideas, and one set of words to clothe them in, and these are always ready at the mouth. So people come faster out of a church when it is almost empty, than when a crowd is at the door". It is therefore probable that these famous essayists would have condemned, and rightly, any tests which purported to test intelligence if success in them depended on mere fluency.

A speaker is deemed fluent if he has command over language and can express his thoughts easily without pause. The profundity of his discourse is not particularly taken into consideration as long as he is talking "sense". The listeners give no thought to the hours or days which the speaker may have needed in order to attain his heights of mellifluence. Similar considerations apply to the case of the fluent writer and his readers. But granted that an apparently fluent speaker or writer may not be innately fluent, it may nevertheless be possible to obtain considerable information if the individual in question is closely observed. It is not difficult to ascertain who can make an impromptu speech or rapidly pen a letter to the press. It is not difficult to pick out those acquaintances who are voluble or taciturn. But merely to label Smith voluble and Brown taciturn does not satisfy the psychologist. One psychologist might hold that there is a relation between volubility or fluency of speech and fluency or quickness in other operations, and that there is a general speed or fluency factor. In that case mental speed would constitute a functional unity so that a person who is quick in one mental process would show a similar quickness in other mental processes, quite apart from the degree of intelligence which he may possess.

[1] Jonathan Swift, *Thoughts on Various Subjects.*

A second psychologist might deny this and hold that Smith's volubility is merely a rather specific verbal ability, a skill in using words which may show his mental capacity in altogether too flattering a light. Has not Healy introduced the term "verbalist" to characterise the individual whose capacity to use language is very much above his ability in other ways?

A third psychologist might identify fluency with intelligence.

A fourth might vigorously oppose such a view and draw an emphatic distinction between those who can think quickly and those who can think well. Such appears also to be the view of Landor the essayist: "Quickness is among the least of the mind's properties and belongs to her lowest estate. The mad often retain it; the liar has it; the cheat has it; we find it on the race-course and at the card table. Education does not give it and reflection takes away from it."

Fortunately the experimental evidence available appears sufficient to enable us to decide between these various hypotheses. The real facts are somewhat complex and are not such as might have been surmised on *a priori* grounds. Thus the experimental approach serves both as the testing furnace of hypotheses made and also as the fount for formulating further hypotheses. The discussion will be made clearer when it is realised that fluency has a noetic aspect and a reproductive aspect. The former has to do with the influence of fluency or speed in tests of "intelligence", the latter with the influence of fluency in tests of "imagination". A simple illustration makes clear the difference between the noetic or eductive aspect on the one hand and the associative or reproductive on the other:

soot : black : : snow :
black : soot : : white :

The finding of the correlative "white" in the first example is eductive and there can be but one answer.

The finding of correlatives like salt, snow, note-paper, etc., in the second example is reproductive.

Fluency in Eduction.—It is expedient to consider the former aspect first. It is obvious that tests of "intelligence" demand a certain power or goodness of response and also a

speed or fluency of response. And the connection between the goodness and the fluency is that of being interchangeable. If the conditions of the case are such as to eliminate the influence of fluency to a large degree, then "g" measures goodness. When—as is most usual—both influences are in play, then "g" measures the efficiency compounded of both. Neither goodness nor fluency constitutes a group factor or produces specific correlation. Thus the view that some persons are on the whole unable to think quickly and yet are quite able to think clearly would seem to be a most grave error.

Fluency in Reproduction.—But this failure to substantiate a fluency factor in its noetic aspect does not mean that a fluency factor does not exist elsewhere. Indeed the research of Hargreaves[1] definitely points to a fluency factor in its reproductive aspect in the sphere of imagination.

The results of Hargreaves obtained corroboration from the experiments of Karve[2] who applied the following seven tests to 33 boys in Standard VI. in one school and 35 girls in Standard VI. in another school in Leeds.

1. *Nouns Test.*—"I want you to write as many nouns as possible in two minutes, beginning with a certain letter. Start immediately after I give you the letter."

The letters P, T, S and H were used.

2. *Unfinished Stories.*—"I am going to give you the beginning of a story and I want you to write the story completely in your own words. You are to write the longest story you can, because the longer the story is the more marks you will get, provided that everything you write has something to do with the story."

Twenty minutes allowed for each.

(*a*) "A boy was so late coming home from school one day that his parents got very anxious. At last he arrived safe though very tired, but with only one boot and no cap, and told them what had happened which was this: when he left school . . ."

(*b*) "A small girl, after her first visit to the Zoo had a very strange dream. She dreamt that . . ."

[1] H. L. Hargreaves, "The 'Faculty' of Imagination", *Brit. Journ. of Psychol.* Monog. Suppl. No. 10, 1927.

[2] B. D. Karve, "An experimental Investigation of 'Fluency' in School Children", M.Ed. Thesis, Univ. of Leeds, 1929.

3. *Controlled Associations.*—Time allowed for each, two minutes.

"Write down as many names as you can of—

(1) Things made of leather.

(2) Things made of wood.

(3) Things that we eat (not drink).

(4) Flowers.

(5) Animals.

(6) Fruits.

(7) Persons you have read about in history.

(8) Towns, cities or villages."

4. *Picture Completion Test.*—"Write down all the things you would put in the picture if you had to complete it."

Four unfinished pictures were used. Three minutes allowed for each.

5. *Predictions.*—"Write down as many different predictions as possible."

For practice and not to be marked: "What might happen if it became unnecessary for people to eat and drink?"

The following four were used and five minutes allowed for each.

(1) "What would happen if everyone could walk and swim at the rate of 100 miles an hour?"

(2) "What would happen if nobody had any teeth?"

(3) "How might our life be different from what it is now if reading and writing had never been invented?"

(4) "What would happen if there was no water on this earth?"

6. *Ink Blots.*—Four blots were selected from Whipple's Test[1] and greatly enlarged copies were made for group use by means of an epidiascope. They were exposed one at a time for two minutes, and the subjects wrote down as many names as were suggested to them by the blots.

7. *Free Association.*—"Write as many different words as you think of." Six minutes allowed.

Karve found the reliability of each of these tests to be satisfactory, *e.g.* in the case of the boys the coefficients of reliability

[1] G. M. Whipple, *loc. cit.* vol. 2, p. 254.

of the seven tests were 0·78, 0·62, 0·78, 0·74, 0·71, 0·84, 0·88 respectively. On eliminating "g", a table of partial inter-correlations was obtained. Then the median value of the tetrad differences was calculated, and also the probable error of the tetrad differences (cf. Chapter 23).

As a result it was concluded that there was a group factor operative in each of the above tests of fluency.

Applications to Psychiatry.—When the present writer[1] applied the Nouns Test to 23 cases of mania and 21 cases of melancholia, the median marks were 17 and 10 respectively. The corresponding marks for the Blots Test being 16 and 8. These results seem to corroborate the general view that usually there is a fluency of associations in mania and a lack of fluency in melancholia, but before drawing definite conclusions it would, of course, be necessary to get corresponding measures from normal subjects drawn from the same *milieu*.

QUESTIONS

1. Discuss the validity as a measure of intelligence of the Binet-Simon Test which expects a child of eleven to give sixty words in three minutes.
2. How would you ascertain if there is any relation between the fluency factor discussed in this chapter and the verbal factor mentioned in Chapter 8?
3. Devise tests which might enable you to ascertain whether the fluency factor is operative even in cases where the verbal factor is almost or entirely absent.

[1] Ll. Wynn Jones, "Individual Differences in Mental Inertia", *Journ. of Nat. Inst. of Industrial Psychol.* vol. 4, No. 5, 1929, p. 287; cf. also *Journ. of Mental Science*, October, 1928.

CHAPTER 11

REACTION TIMES

FREE ASSOCIATION (STRESSING FLUENCY)

Experiment 1.

MATERIALS.—Kent-Rosanoff list of 100 words (cf. G. H. Kent and A. J. Rosanoff. A study of association in insanity, reprinted from *Amer. Journ. of Insanity*, vol. 67, 1910).

The Self-Recording Chronograph, and the Tenth of a Second

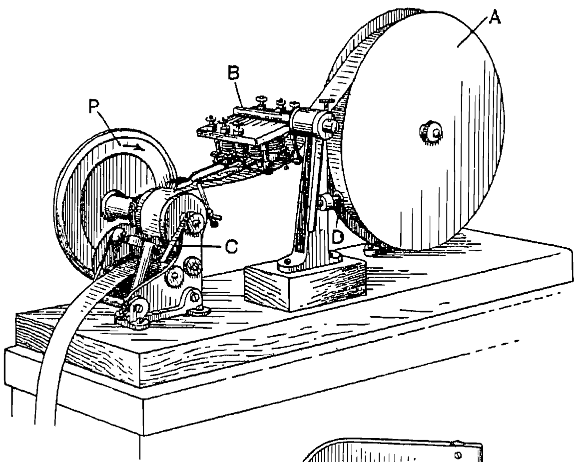

FIG. 18.

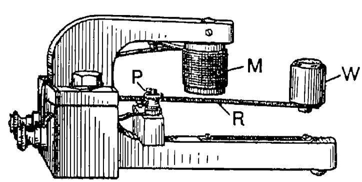

FIG. 19.

Vibrator with Motor obtainable from Messrs. C. F. Palmer (London), Ltd., 63A Effra Road, Brixton, London, S.W.2.

DIRECTIONS.—It is advisable to commence with the following experiment which pays regard to the qualitative aspect: S, with his back to E, is requested to respond as quickly as possible to each word read by E by saying the first word that occurs to him

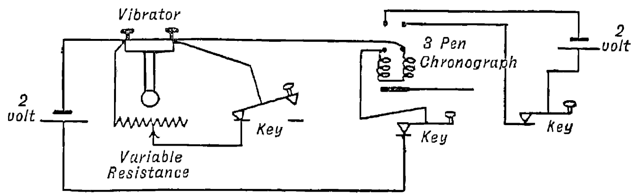

FIG. 20.

other than the stimulus word. E starts the stop-watch as he finishes reading the word and stops it when S begins to respond. The first ten words in the Kent-Rosanoff list may be used. The results should be entered under the following headings in four colours:

Stimulus Word Reaction Word Reaction Time Introspection

After each response S attempts to trace how the reaction word arose, and his replies are entered under the heading "Introspection". For the second ten words in the Kent-Rosanoff list E and S change places. Afterwards E and S should attempt to classify the responses according to their nature.

After this preliminary work the pair should now set up the Self-Recording Chronograph which is suitable for the more quantitative study of reactions. The writer recommends the following procedure: The 100 words of the Kent-Rosanoff list should be divided into ten lists of ten words; each list of ten words should appear in column on a piece of cardboard, each word being printed in capitals about a quarter of an inch high, and a space of one inch between each word. Actually eleven words appear on each sheet of cardboard, as each list is preceded by a word not included in the Kent-Rosanoff list. The Kent-Rosanoff list is as follows:

(1) Table	(11) Black	(21) Sweet	(31) Rough
(2) Dark	(12) Mutton	(22) Whistle	(32) Citizen
(3) Music	(13) Comfort	(23) Woman	(33) Foot
(4) Sickness	(14) Hand	(24) Cold	(34) Spider
(5) Man	(15) Short	(25) Slow	(35) Needle
(6) Deep	(16) Fruit	(26) Wish	(36) Red
(7) Soft	(17) Butterfly	(27) River	(37) Sleep
(8) Eating	(18) Smooth	(28) White	(38) Anger
(9) Mountain	(19) Command	(29) Beautiful	(39) Carpet
(10) House	(20) Chair	(30) Window	(40) Girl

(41) High	(51) Stem	(61) Memory
(42) Working	(52) Lamp	(62) Sheep
(43) Sour	(53) Dream	(63) Bath
(44) Earth	(54) Yellow	(64) Cottage
(45) Trouble	(55) Bread	(65) Swift
(46) Soldier	(56) Justice	(66) Blue
(47) Cabbage	(57) Boy	(67) Hungry
(48) Hard	(58) Light	(68) Priest
(49) Eagle	(59) Health	(69) Ocean
(50) Stomach	(60) Bible	(70) Head

(71) Stove	(81) Butter	(91) Moon
(72) Long	(82) Doctor	(92) Scissors
(73) Religion	(83) Loud	(93) Quiet
(74) Whisky	(84) Thief	(94) Green
(75) Child	(85) Lion	(95) Salt
(76) Bitter	(86) Joy	(96) Street
(77) Hammer	(87) Bed	(97) King
(78) Thirsty	(88) Heavy	(98) Cheese
(79) City	(89) Tobacco	(99) Blossom
(80) Square	(90) Baby	(100) Afraid

The following words are suggested as suitable for starting each list: garden, wood, coat, pebble, mouse, pity, sky, grass, rain, swan. E now gives S the third list of eleven words, face downwards, and at a signal S turns it over and gives as quickly as possible the first word that occurs to him on seeing the first stimulus word. Without pause he does the same to the second

stimulus word and so on. Each reaction is registered by E by pressing the key. E and S change places alternately until the lists have been completed.

CONSTRAINED ASSOCIATION (STRESSING RETENTIVITY)

Experiment 2.

DIRECTIONS.—Lists of easy French words are arranged in a column as in Experiment 1 and S is instructed to give the English equivalents as quickly as possible. At each response E taps on the key as before. For the next list E and S interchange places. The following are two lists which may prove suitable:

First list: matin, quand, votre, cher, école, pourquoi, frère, jour, fenêtre, arbre, beaucoup.

Second list: ici, maison, jamais, soleil, quatre, chien, très, maintenant, cheval, autre, oui.

Henmon[1] found that such words can be correctly translated by more than 90 per cent of pupils who have studied French for one year. Should S be unable to respond to any word, he should say "No" and immediately pass on to the next. The median reaction time should be ascertained.

CONSTRAINED ASSOCIATION (STRESSING NOEGENESIS)

Experiment 3.

DIRECTIONS.—As before, lists of suitable words should be prepared, eleven on each list, and S is instructed to give the opposites to them as quickly as possible. At each response E taps the key. For the next list E and S interchange places. Whenever S fails to respond he should say "No" and proceed to the next word and E notes at which points these failures occur so as to leave them out of the calculations when necessary. The median reaction time is again noted.

THE SIMPLE REACTION

Experiment 4.

DIRECTIONS.—The gearing of the motor in this experiment should be arranged to furnish a speed such that the distance

[1] V. A. C. Henmon, "Standardised Vocabulary and Sentence Tests in French", *Journ. of Educ. Research,* vol. 3, 1931.

travelled in one-tenth of a second by the chronograph paper is appreciable. E's key operates one magnet and S's key operates another.

(*a*) "NATURAL" REACTION.—At the word "Ready" S sits with closed eyes, with his finger on the key. E then says "Now" and after an interval of about two seconds taps sharply upon his own key. S's task is to press his key as soon after E as possible. S then writes his introspection with special reference to the composition of his motive and the direction of his attention. After ten experiments E and S change places.

(*b*) "SENSORIAL" REACTION.—At the word "Ready" S closes his eyes and places his finger on the key as before. But he is now told to attend to the sound of E's key, and as soon as he hears it, he is to press his key as quickly as possible. He is to pay attention to the sound and not to the pressing of his key as that will look after itself. After ten experiments E and S change places.

(*c*) "MUSCULAR" REACTION.—At the word "Ready" S closes his eyes and places his finger on the key. But he is now told to pay attention to the pressing of his key as quickly as possible when the sound comes. He is to attend to the intended movement and not to the sound. After ten experiments E and S change places.

When the experiments are completed, the average reaction time for the "natural", "sensorial" and "muscular" reactions should be calculated, and also in each case the mean variation. But in order to get a comparison between "sensorial" and "muscular" reaction times, it is well to disregard the first series of tests and then to allow for practice by the following procedure: start with 25 "sensorial" trials, then 50 "muscular" trials, and finally 25 more "sensorial".

THE VERNIER CHRONOSCOPE

Supplementary Experiment

MATERIALS.—Two pendulum bobs, two lengths of stout silk threads. Cast-iron base and upright, with a clamp for attaching a cross-bar to which the bobs are suspended. Two clips to serve

as reaction keys, the reaction consisting in releasing a bob from the jaws of a clip to which it is held by a thread. Stop-watch.

Directions.—E places one pendulum in position and releases it by pressing the clip and counts the number of full swings in one minute. The length of the thread is adjusted until there are 75 swings to the minute. Hence the time for a single swing is 0·8 sec. E then repeats the procedure with the other pendulum until there are 77 full swings to the minute, so that the time of a single swing is 0·78 sec. approximately. It is then clear that the longer pendulum will make 39 full swings while the shorter is making 40. The plan of the experiment will now become clear. After the word "Ready" S places his finger on his key and closes his eyes. E then gives the signal "Now" and about two seconds later presses his own key which releases the longer pendulum. S then presses his key in reply, and in so doing releases the shorter pendulum. E counts the number of swings of the longer pendulum from the starting of the shorter pendulum to the moment of complete coincidence of the four threads. This number gives the reaction time in fiftieths of a second as can easily be verified if the principle of the Vernier is applied. By such simple and inexpensive means then it will be possible to repeat the experiments on "natural", "sensorial" and "muscular" reactions described in the last paragraph with a unit of measurement of a fiftieth of a second. Full directions are given in E. B. Titchener, *Experimental Psychology*, vol.1, Part 1, p. 117, and vol. 1, Part 2, p. 212. The present author does not therefore consider it necessary to recommend for this course a reference to Hipp's chronoscope which was designed to measure as short a period as a thousandth of a second.

DISCUSSION

"The reaction time of an individual is the interval that elapses between the exhibition or application of a stimulus to him, and his response to it in a prescribed manner."[1] Its study was taken in hand by astronomers, under the title of "the personal equation", long before psychology became an experimental science.

[1] C. S. Myers, *Textbook of Experimental Psychology*, p. 125.

Maskelyne dismissed his assistant Kinnebrook at Greenwich in 1796 because his observations yielded times of stellar transits which were about eight-tenths of a second later than those obtained by Maskelyne himself. It is interesting to note that when this incident became known to Bessel twenty years later, it helped him to discover not only the personal equation but also its variability.

Eventually the "eye and ear" method of Bradley, which involved the observer's reaction time, was discarded in favour of electrical methods. It was early realised that the study of reaction times was essentially a psychological task. Reaction time was found to vary with the sense organ stimulated, with the intensity, duration or speed of the stimulus, with the directions given to the subject, or with his attitude.

It was customary to distinguish between (*a*) simple reactions and (*b*) complex or compound or composite reactions.

(*a*) Simple Reactions.—L. Lange[1] in 1888 found that the "sensorial" reaction is, on the average, about a tenth of a second longer than the "muscular", while the average time for the "natural" reaction is intermediate between the other two.

Lange's Attention Theory was disputed by some investigators. Thus Cattell and his pupils stressed the part played by practice. After practice, the sensorial reaction was only a little longer than the muscular. There is also Baldwin's Type Theory which maintains that individuals are innately sensorial, muscular, or natural in type. For each individual, the quickest reaction would be that which accords with his innate disposition. With so many variables involved it would be a formidable undertaking to decide between these theories, and besides, they are no longer questions of the day.

The student who performs Experiment 4 will probably be inclined to agree with Lange's view, but it is obvious that Experiment 4 does not enable him to assess Cattell's theory or Baldwin's theory. Simple reaction time is further dependent on the structure of the sense organ stimulated. The following

[1] L. Lange, *Wundts Philos. Studien*, 4, 1888. Titchener (*Exp. Psychol.* vol. 1, Part 2, p. 214) points out that Lange was anticipated by S. Orschansky in the *Neurol. Centralbatt*, No. 12, 1887.

table[1] shows that visual reactions are slower than auditory or tactual, and the reactions for smell, taste and especially pain are slower still:

Stimulus	Muscular Reaction	Sensorial Reaction	
Light .	175 σ	270 σ	$\sigma = \frac{1}{1000}$ sec.
Sound .	125 σ	220 σ	
Touch .	110 σ	210 σ	

(*b*) COMPLEX REACTIONS.—Here central processes, both cognitive and volitional, may cause the reactions to be much slower than in the case of the so-called simple reactions. There are various ways by which the complexity of the reaction may be increased.

(1) *Recognitive Reactions.*—Instead of measuring S's sensorial reaction to a visual stimulus, *e.g.* a colour, he may be instructed not to react until he has *recognised* which colour it is. Then the reaction time is somewhat slower.

(2) *Discriminative Reactions.*—This may be complicated by presenting to S one or other of a number of known stimuli and instructing him not to react until he has *discriminated* the presented simulus from the other possible stimuli.

(3) *Choice Reactions.*—One form, devised by Donders, is to present either of two stimuli, to only one of which S is instructed to react.

A more complex form due to the same investigator is to employ two stimuli and to instruct S to react with the right hand key if one stimulus appears, and with the left hand key if the other.

Still another form is the ingenious arrangement of Henmon.[2] Two stimuli A and B are presented side by side. S is instructed to react with his right hand if A appear on the right of B, and with his left hand if it appear on the left of B.

ASSOCIATIVE REACTIONS.—Here the association may be "free" as in Experiment 1, or it may be "constrained" as in

[1] C. S. Myers, *Textbook of Experimental Psychology*, p. 126.

[2] V. A. C. Henmon, "The Time of Perception as a Measure of Differences in Sensations", *Arch. of Phil., Psychol. and Sci. Methods*, No. 8, 1906.

Experiments 2 and 3, or it may be "partly constrained" when S has a choice of answers, such as to give a specific instance of some generic word.

REACTION TIMES AND SPEARMAN'S PRINCIPLES

It is one step to recognise that one reaction is "simple" and another "complex". It is a further step to ascertain the nature of the complexity by the aid of fundamental principles. Here the Law of Retentivity and those of Eduction are potent aids, and the evidence suggests that regard must also be given to a factor of Fluency. To carry out the analysis, the experiments described in this chapter may be recommended. In each experiment it will be noted that the instructions require the reaction words to be given as quickly as possible. This must be kept in mind in Experiment 1, as the Kent-Rosanoff procedure did not emphasise this.

Experiment 1: *Free Association*

Using the given list of 100 words, Kent and Rosanoff tested one thousand normal subjects. They were of both sexes, of ages ranging from eight to eighty years, they had various degrees of mental capacity and education, lived in widely separated localities and included persons following various occupations.

To each of the 100 words, S was directed to react by the first word of which it made him think. For each stimulus word a frequency table was compiled which exhibited the frequency of every word elicited in response. Thus to the word "table", the leading frequencies were: chair 267, wood 76, furniture 75, eat 63, cloth 57. But there were many responses such as basket, ink, oak, which were only given by one individual out of the thousand. Kent and Rosanoff eventually took no account of the reaction times. The aim of their investigation was to study pathological variations in reactions. With this in view they regarded a reaction as *common* if it appeared in their tables. If not, it was regarded as an *individual* reaction, except that a reaction was classed as *doubtful* if it was a grammatical variant of a word found in the tables. When analysing the record of each of their normal subjects, it was removed from the mass of

material which formed the tables, and the reactions of the remaining 999 subjects served as a standard of comparison.

Kent and Rosanoff pointed out that a small number of their thousand subjects exhibited various types of abnormal reactions. On the other hand, there must be a certain number of words which, although they do not appear in their tables, must be regarded as normal reactions. Reactions regarded as pathological were: derivatives of stimulus words; non-specific reactions, *i.e.* words such as article, good, great, necessary, person, which are so widely applicable as to serve as reactions to almost any stimulus word; sound reactions; word complements; perseverative reactions, *i.e.* associations to a preceding stimulus or reaction, or repetitions of preceding stimuli or reactions.

The normal subjects gave, on the average, 6·8 per cent individual and 91·7 per cent common reactions. But the range of variations was rather wide, a considerable number of subjects giving no individual reactions at all while a few gave over 30 per cent. On the other hand, 247 insane subjects gave, on the average, 26·8 per cent individual and only 70·7 per cent common reactions.

SPEED IN ASSOCIATIVE REPRODUCTION

Instead of stressing the reaction word rather than the reaction time the present writer,[1] using Palmer's apparatus, employed the opposite procedure. Boys from Forms 3A and 5A of the City of Leeds School[2] were tested individually. S was given a cardboard sheet with eleven common English words printed in column, and held face downwards. At a signal he turned over the card and gave in order and as quickly as possible the first word that occurred to him on seeing each stimulus word. Simultaneously with each response E pressed his key. The reaction to the first word is therefore represented on the moving tape of the chronoscope and serves as the point from which the reaction-time for the second stimulus word is measured. Actually this was not done as the series merely served as preliminary practice. There followed then the First

[1] Ll. Wynn Jones, *A Study of Speed in Associative Reproduction*, Rep. of Brit. Assoc., 1931.

[2] My thanks are due to the Headmaster, Mr. Worts, for valuable help.

Association Test carried out in a similar way and recorded. It consisted of the eleven words: wood, black, mutton, comfort, hand, short, fruit, butterfly, smooth, command, chair.

S was now given a card with eleven common French words printed in column. At a signal, as before, the card was turned over and S translated each word in order into English as quickly as possible. Simultaneously with each response E pressed his key. After this preliminary practice there followed the First Translation Test consisting of the following eleven words: matin, quand, votre, cher, école, pourquoi, frère, jour, fenêtre, arbre, beaucoup.

A fortnight later two similar tests were given, namely, the Second Association Test consisting of the words: mouse, high, working, sour, earth, trouble, soldier, cabbage, hard, eagle, stomach; and the Second Translation Test consisting of the words: ici, maison, jamais, soleil, quatre, chien, très, maintenant, cheval, autre, oui. Such words, according to Henmon, are known to over 90 per cent of pupils after studying French for only one year. Form 3A had studied French for 1½ and Form 5A for 3½ years. It will be seen that the difference in speed between the two Forms in the Association Tests is not at all pronounced. But in the Translation Tests the speed in the lower Form is on the average about 50 per cent slower than that of the upper which has studied French for two years longer. It is also evident that both in the association and in the translation test one pupil in a class may respond twice, three times, or even four times as quickly as another.

	Reaction Times (in seconds)		
	Quickest	Median	Slowest
Form 3A 1st Association Test .	1·10	1·77	4·00
2nd Association Test .	1·20	1·93	3·50
Form 5A 1st Association Test .	0·97	1·70	3·55
2nd Association Test .	0·95	1·62	3·17
Form 3A 1st Translation Test .	0·60	0·92	1·52
2nd Translation Test .	0·77	1·17	2·97
Form 5A 1st Translation Test .	0·40	0·62	0·92
2nd Translation Test .	0·57	0·80	1·20

Cattell reports that one of his observers took on the average 0·152 sec. to translate familiar German words into English, and 0·258 sec. in the reverse direction. But his results were obtained by the usual discrete method of using a lip-key and therefore cannot be compared with the present results.

The average age of the 35 boys from 3A was 13½ years and of the 23 boys from 5A 15½ years. The class mark in French, the Easter examination mark in French, the mark for all school subjects pooled, and the score in the Cattell Group Intelligence Scale were also available for each boy. The speed of translating well-known French words into English showed appreciable correlation with the class marks in French, with the examination marks in French, and also with the marks for all school subjects pooled. As would be expected the Cattell test also correlated with this pool, but the slight positive correlation between intelligence and speed of translation in the lower form has entirely disappeared and is actually negative in the upper. There is therefore no reason why pupils with even less than average intelligence should not attain high proficiency in a modern language after remaining the normal period at a secondary school. They may not be gifted enough to pursue higher language studies but they may be eminently suitable for commercial positions in virtue of their ability for retaining foreign words and their possession of other suitable traits.

QUESTIONS

1. Discuss critically the classification of reactions into simple and composite reactions.
2. Compare the reaction time of your subject in a free association experiment with that in typical forms of the constrained association experiment. How can you account for the differences obtained?

References

1. E. B. Titchener, *Experimental Psychology*, vol. 2, Parts I and 2, chap. 3, The Reaction Experiment, 1905.
2. V. A. C. Henmon, "The Time of Perception as a Measure of Differences in Sensations", *Arch. of Phil., Psychol. and Sci. Methods*, vol. 8, 1906.

3. C. S. Myers, *Experimental Psychology*. chap. 11, Reaction Times, 1911.
4. J. McKeen Cattell, "Some Psychological Experiments", *Science*, January 1 and 8, 1926.
5. E. G. Boring, *A History of Experimental Psychology*, chap. 8, The Personal Equation, 1929.
6. H. E. Garrett, *Great Experiments in Psychology*, chap. 9, Cattell's Studies in the Measurement of Reaction Time, 1930.

CHAPTER 12

PROGRESS IN LEARNING AND TRANSFER OF TRAINING

FIGURE SPOTTING TEST

DIRECTIONS.—S is instructed to write A under 1, B under 2, and so on, and when all the letters of the alphabet are used, to continue with A, B, C, . . . again until there is a letter under every number, in the following table. The figures are covered and E gives the signal "Ready", then two seconds later the signal "Go", at the same time starting the stop-watch which is stopped when S has finished the task. At uniform intervals the same task is repeated by S at least twelve times. S should write an introspective report on his progress and should plot the learning curve, the number of attempts along the horizontal axis and the times along the vertical. A composite curve may be drawn by pooling the results of all the Ss.

A sufficient number of copies of the table should be prepared beforehand.

6	35	30	19	23	3
26	32	10	28	5	17
21	1	15	25	12	36
11	18	29	8	22	14
4	31	24	33	27	16
7	34	2	13	20	9

DISCUSSION

The above test is only one of countless others which may be employed in order to study the factors operative in learning. These factors obviously vary with the test, and the shape of the

learning curve depends on the nature of the task.[1] The initial progress is relatively rapid in some and relatively slow in others. In most curves, though not in all, plateaux are found, that is, the curve at a certain stage is flat and further practice does not for a time cause improvement. Amongst the possible causes of these plateaux may be cited:

(1) Decrease in efficiency due to objective fatigue (cf. Chapter 19).

(2) Lack of motive for working at full pressure. Flugel's curves were obtained from the results of S's working at full pressure, and as the incentives were powerful they tended to eliminate the plateaux (cf. Chapter 19).

(3) The formation of the simpler habits demanded by the task is naturally facilitated by repetition and then a stage arrives when S confidently attempts the formation of more complex habits, but before he attains proficiency it is not surprising if there occurs a plateau or even a set-back (cf. the mastery of typewriting or piano playing). It is at least certain that many plateaux can be eliminated by the avoidance of fatigue, by the employment of suitable incentives and by efficient methods of learning.[2] But finally, continued attempts at perfection will cause the subject to approach his physiological limit.

THE TRANSFER OF LEARNING

After practising the Figure Spotting Test for many days it may be asked whether the training would make S efficient in a Letter Spotting Test, that is, in a test which necessitated the writing of the figures 1, 2, etc., under the letters A, B, etc., which were placed in haphazard order. It may be objected that such a test would be too "picayunish" to exert any natural appeal and, moreover, that the transfer of efficiency, if present, is only carried over to a more or less similar activity. In fact it

[1] Cf. R. Pintner, *Educational Psychology*, 1929, chap. 9, The Learning Curve; and D. Starch, *Educational Psychology*, 1927, chap. 11, The Rate and Progress of Learning.

[2] Cf. E. Healey, "The Plateau: A Study of Periods of Arrested Progress in Motor Learning." M.Ed. Thesis, Univ. of Leeds, 1934.

must be acknowledged that an experiment on transfer which would be really worth while is likely to require more time than can be allowed for in this book. In the past the doctrine of transfer of training or formal discipline was concerned with the general training of the so-called faculties of the mind which was supposed to result from the study of a particular subject. Thus the study of Latin was supposed to train the "powers of reasoning" or to develop "judgment", and should therefore be studied by all potential legislators. Equally sweeping claims have been made for mathematics, science, modern languages, hand-work, or what not. It is not asserted here that these subjects have no effect on the individual apart from their utilitarian value, but the above claims were in the nature of dogmatic statements which did not rest on facts. When the doctrine of formal discipline was subjected to a rigid experimental test by Thorndike and others its claims for consideration were reduced to a minimum. A recent monumental investigation by Thorndike and his collaborators enabled him to write these words: "The expectation of any large difference in general improvement of the mind from one study rather than another seems doomed to disappointment. The chief reason why good thinkers seem superficially to have been made such by having taken certain school studies, is that good thinkers have taken such studies, becoming better by the inherent tendency of the good to gain more than the poor from any study. When the good thinkers studied Greek and Latin, these studies seemed to make good thinking. Now that the good thinkers study Physics and Trigonometry, these seem to make good thinkers. If the abler pupils should all study Physical Education and Dramatic Art, these subjects would seem to make good thinkers. These were, indeed, a large function of the program of studies for the best thinkers the world has produced, the Athenian Greeks. After positive correlation of gain with initial ability is allowed for, the balance in favour of any study is certainly not large. Disciplinary values may be real and deserve weight in the curriculum, but the weights should be reasonable." [1]

[1] E. L. Thorndike, "Mental Discipline in High School Studies", *Journ. of Educ. Psychol.* vol. 15, 1924, p. 98.

DISCIPLINE, UTILITY, CULTURE.—Divergent views regarding the aims of education will always be held, for educational writers differ widely in nurture. The first may have been educated privately, the second at Eton and Christ Church, the third entered the lists through the parish school and the coal mine, the fourth may be a philosopher, the fifth an applied chemist, the sixth an aesthetician. When, further, account is taken of their wide differences in intellectual and temperamental equipment, it would indeed be a miracle if they all stressed "disciplinary", "utilitarian" and "cultural" values alike. We are not here concerned with the aims of education, but would nevertheless insist that the divergence of views is partly a question of terms. It can safely be said that terms like "cultural value" are at present hopelessly inexact for scientific usage. Thus cultural training is often praised at the expense of vocational training. Such loose thinking in pitting the one against the other may be exposed by pointing out that a teacher, clergyman, physician, or lawyer who has neglected his vocational training in psychology must surely lack culture. A comprehensive definition of culture which may avoid this antithesis is given by Brierley,[1] namely, "an integration of knowledge and experience which helps an individual to lead a reasonably full and free intellectual and emotional life and fits him to play a contributory part in the social community; in short, an orientation towards life which makes for personal happiness and social usefulness." It follows from such a view that it would be difficult, if not impossible, to estimate the "disciplinary", "utilitarian" or "cultural" value of a school subject. This difficulty was experienced by Starch[2] when he requested 58 principals and teachers to assume the disciplinary value of the first year of high-school English as taught in the average way to be 10, and that all other values be estimated in terms of this assumption. For many obvious reasons Starch did not attach much importance to his tabulation, *e.g.* teachers would tend to rank their own specialities high. It is

[1] W. B. Brierley, *The Cultural Value of Science in Adult Education*, Rep. of Brit. Assoc. Sect. L, 1933.

[2] D. Starch, *Educational Psychology*, 1920, p. 219.

therefore necessary to know what subjects were taught by the teachers concerned. The estimates obtained varied widely from teacher to teacher, nevertheless there was a tendency to cluster round a mean.

At Easter 1933 I repeated Starch's experiment. Over 90 secondary school teachers and training college lecturers attended a Refresher Course and 25 volunteered to give estimates. In several respects the results agreed with those of Starch. Thus the highest disciplinary value was assigned to Geometry. High utility values were given to English, Arithmetic, Cooking, Needlework and Carpentry. English and Music had the highest cultural values. But, on the other hand, startling variations were exhibited in the estimates. Thus the disciplinary value of French, Latin or Physical Training varied from 2 to 40, of Cooking between 0 and 20; the utility value of Chemistry from 0 to 40, of History from 2 to 30, and of Cooking from 8 to 50; the cultural value of Latin from 0 to 30, and of Geography or French from 2 to 40.

Obviously no importance should be attached to these figures; they only serve to show that terms like "cultural value" have no fixed meanings. In fact, the experiment may be regarded as meaningless for, in the words of A. E. Hirst, who took part in it, "the three values may be so closely related and interdependent that it is impossible to regard a subject as having, say, greater or lesser utilitarian than disciplinary or cultural value. It is also possible to make any one of these subjects a 'core' subject."

Theories of Transfer.—What factors determine the amount of transfer? THORNDIKE'S THEORY OF COMMON ELEMENTS is based on the doctrine that the learning of a task consists of the establishment of specific connections between various specific elements, or that the learning of a task is essentially the establishing of minute reactions to a complex situation. Now, if a second task contains elements that activate some of the same reactions so that certain elements are common to the two tasks, there will be transfer and the amount of transfer from one task to the other will be proportional to the amount of identity. Thorndike considers the identical elements to be

specific connections in the nervous system which are common to the tasks or functions. As indicated elsewhere in this book such a neurological reference may not be very helpful to the psychologist, and still less so to the educator. But it is clear from Thorndike's exposition that his common elements have a very wide significance. Thus he writes:[1] "Chief amongst such identical elements of practical importance in education are associations including ideas about aims and ideas of method and general principles. . . ." These elements, then, may be elements of—

(1) *Content*, *e.g.* the study of Latin helps that of French in so far as the words are similar, such as "terre" from "terra".

(2) *Method*, *e.g.* learning a stanza as a whole rather than line by line.

(3) *Ideal*, *e.g.* the ideal of neatness cultivated in connection with one special subject may be made to irradiate into other activities.

JUDD'S THEORY OF GENERALISATION OF EXPERIENCE explains the transfer in terms of the recognition of application of an experience obtained in one connection to other connections. It emphasises the conscious recognition of the identical elements in as many situations as possible.[2]

There is not necessarily any opposition between the two theories, and each has helped to elucidate the effects of transfer. But an alternative explanation of the facts of transfer is possible by the use of SPEARMAN'S TWO FACTOR THEORY AND HIS DOCTRINE OF NOEGENESIS. Here "common relations" would supply the key to the solution of many a problem in transfer.

THE METHOD OF EQUIVALENT GROUPS

The technique commonly employed in experiments on transfer may be made clear by a concrete example. Suppose we wish to ascertain how a month's practice in learning poetry would affect the ability to learn prose. The first step is to pick two equivalent groups, A and B. Prose Test X is given, and on the

[1] E. L. Thorndike, *Educational Psychology*, vol. 2, 1914, p. 259.

[2] C. H. Judd, *The Psychology of High School Subjects*, 1915, p. 413.

result of this test the groups A and B are picked in such a way that for every member of group A there is a corresponding member of group B of the same ability for learning prose. Thus the two groups will have the same average ability for learning prose and the same variability with reference to this average. For an exact experiment it would be necessary to equate the groups with respect to intelligence, schooling, age, race and sex and other variables as well. But, in practice, it is obvious that this could only approximately be effected. Group A, the Practice Group, is then given practice in learning poetry every day for a month, while Group B, the Control Group, gets no practice. At the end of the month both groups take another Prose Test Y.

Group A: Prose Test X—Practice in Poetry—Prose Test Y.
Group B: Prose Test X—No Practice in Poetry—Prose Test Y.

Assuming that the Prose Test Y is equal in difficulty to the Prose Test X, it is possible that both groups will improve in Test Y partly as a result of practice. But if Group A has a higher average score than Group B in Test Y, it is possible to determine whether the gain is significant (cf. Chapter 23). If there are no grounds for assuming that the two prose tests are of equal difficulty, it is more exact to employ four groups A, B, C and D. Groups A and C get practice in learning poetry, B and D get no practice. The following scheme will make the procedure obvious.

Group A: Prose Test X—Practice in Poetry—Prose Test Y.
Group B: Prose Test X—No Practice in Poetry—Prose Test Y.
Group C: Prose Test Y—Practice in Poetry—Prose Test X.
Group D: Prose Test Y—No Practice in Poetry—Prose Test X.

As in the former case, the four groups have been matched as nearly as possible. The gain of A over B in Test Y, and the gain of C over D in Test X, can now be averaged.[1]

This chapter will be concluded by a brief account of three experimental investigations which connect the problem of

[1] Cf. G. H. Thomson, *Instinct, Intelligence and Character*, 1924, p. 137.

transfer with educational methods, namely, that of Meredith in the teaching of science, that of Hughes in the teaching of arithmetic, and that of Gurney in the teaching of modern languages.

CONSCIOUSNESS OF METHOD AS A MEANS OF TRANSFER.—A recent investigation by Meredith [1] suggests that consciousness of method may be an important agency in the transfer of training. Three groups of boys of approximately equal average intelligence were chosen. Group A was the control group. Group B received incidental practice in defining scientific terms. Group C were *trained* in definition and there was an explicit reference to the form which a correct definition should take. The transfer to be investigated was from the definition of scientific terms to the definition of ordinary words in daily use. When the training was completed all the boys were tested again, and it was found that Group C had benefited by its training in defining scientific terms, whilst Group A and Group B did no better than before. Meredith's results suggest that the methods of science when consciously recognised by the pupils may become ideals which function generally. This is a very different doctrine from that which would assume that the formal discipline, whether in Latin or in science, improved judgment or intellect or what not. In agreement are the conclusions of Whipple [2] when he recommends that teachers should arrange the work of their pupils so as to facilitate the conscious recognition by the pupils of the methods by which mental work is done. Whipple also points out that bright children may be expected to surpass dull children in the amount of transfer gained from specific training. This is what would be expected in view of Strasheim's finding that the higher stage of relation-finding is that where the relations are "abstracted" and have become concepts which are effective for universal use.[3]

[1] G. P. Meredith, "An Investigation of Some Aspects of the Problem of Transfer of Training", M.Ed. Thesis, Univ. of Leeds, 1926. Cf. also, G. P. Meredith, "Consciousness of Method as a Means of Transfer of Training", *Forum of Education*, vol. 5, No. 1, February 1927.

[2] G. M. Whipple, "The Transfer of Training", *27th Year Book of the Nat. Soc. for the Study of Educ.* Part 2, 1928.

[3] J. J. Strasheim, *A New Method of Mental Testing*, 1926, p. 141.

The Drill Value of Number Games.—Another study which has a bearing on the problem of transfer is the important investigation of Hughes. Here the drill value of number games was approached from two standpoints. Hughes formulated six important criteria which good number games should satisfy, and then compared by experiment the drill value of good number games with that of conventional drill methods in arithmetic. Using Ballard's tests of addition, two equal groups were selected, the subjects being Standard II. boys. One group practised addition by drill methods while the other played number games. Fifteen daily periods of fifteen minutes' duration were given, the experiment lasting just over three weeks. Both groups were tested after 10 practices and also at the end. After 10 practices the drill group improved 25·6 per cent and the games group 17·3 per cent, but after 15 practices the corresponding figures were 34·3 per cent and 32·1 per cent. At the conclusion of the testing then, there is no significant difference between the two groups in spite of the facts that the tests favoured the drill group, that the drill given was as good as could be devised, and that the games were only played fifteen times. Hughes therefore suggests that these games are superior to much of the practice as carried on in school. It is obvious that interest in winning the game is a potent factor to account for the improvement shown by the games group, but Hughes[1] points out that under the conditions of the test, this interest was shown also by the drill group as they guessed they were taking part in a competition. Under ordinary school conditions, however, the advantage from the point of view of interest would lie with the games.

THE VALUE OF SILENT READING IN THE TEACHING OF MODERN LANGUAGES

Methods of investigating the value of practice in silent reading in the teaching of English have been demonstrated by

[1] A. G. Hughes, "The Practice Value of Games in Arithmetical Operations", M.Ed. Thesis, Univ. of Leeds, 1922. Cf. also A. G. Hughes, "An Investigation into the Drill Value of Number Games", *Educational Research*, vol. 6, No. 1, September 1927.

many American investigators[1] and by West[2] in India. Gurney[3] appears to be the first to apply such methods to modern language teaching under ordinary English school conditions which inevitably demand some work of the drill and class teaching type. The power of writing a foreign language, the power of speaking it, or the power of understanding it when spoken by others are not activities which are practised by the majority of pupils in real life but, as Gurney points out, the power of *reading* a foreign language may very well be. Silent reading practice was found to be superior to the traditional method of translation for enabling the average pupil to attain reading power and, from the point of view of school organisation, this power may be developed with a comparatively small expenditure of time. Even when the translation trained pupil showed a high reading speed, his comprehension of the narrative was usually not up to the standard attained by silent reading practice. Such practice when given to pupils towards the end of their school course invariably produced considerable increases of speed with increased comprehension, whether the pupils were of high or low linguistic ability. In earlier stages, silent reading proved superior in producing reading power to the study of textbooks of "French Course" type even when these were supplemented by a certain amount of translation of continuous narrative.

A group of 103 boys at the School Certificate stage varied in speed at one of Gurney's tests from 225 to 73 words per minute, the median speed being 127. All had been given a four-, and in some cases a five-years' course in French. On the other hand, a group of third-year boys, after silent reading practice during a total of eighteen lesson periods spread over five months during which they covered about 23,000 words of narrative material, showed an average increase of speed of about 96 per cent, namely, from 76 to 150 words per minute. Seven students who had taken their B.A. Pass Degree in French less

[1] Cf. J. A. O'Brien, *Reading: Its Psychology and Pedagogy*, 1926.

[2] M. West, *Learning to Read a Foreign Language*, 1926.

[3] D. Gurney, "An Experimental Enquiry into the Value of Silent Reading in the Teaching of Modern Languages in Schools", M.Ed. Thesis, Univ. of Leeds, 1931.

than a year before had a median speed of 236, which is approximately the speed of the average 8th Grade pupil in America in reading English when tested by some of the standardised silent reading tests.

The upshot would seem to be that the average pupil after studying French for four or five years cannot be considered to be an efficient reader capable of dealing with ordinary French texts for information or pleasure. In the words of Gurney: "It does not seem to be yet recognised that in order that the pupil may learn to read French, he must be given practice in reading it. We are, as far as this skill is concerned, still inclined to trust to 'indirect' methods."

QUESTIONS

1. What exactly is meant by the term "formal" in the doctrine of "formal Training"?
2. What is meant by the term "formal" in the phrase "formal sense training" as applied to the procedure of Madame Montessori?
3. Can the "form" taken in the sense of "act" be developed independently of the "material" taken in the sense of "content"?
4. Formalists have claimed that training of the "form" would be effective universally, but just the same claim might be made for "material" training. Consider the experimental evidence bearing on this statement.
5. Examine critically the doctrine that "the power to judge is increased by judging".

   (For Questions 1-5 cf. C. Spearman, "Qualified and Unqualified Formal Training", *Journ. of Exp. Pedagogy*, vol. 2, No. 4, 1914.)
6. Give a brief résumé of the results of experiments on transfer with reference to Reaction Time, Perception and Discrimination, Sensori-Motor Association, Memory, and Cross Education (cf. D. Starch, *Educational Psychology*).
7. Give a summary of Thorndike's methods and results in his researches on transfer (cf. *Journ. of Educ. Psychol.* vol. 15, pp. 1-12, 83-89; vol. 18, pp. 377-404).

References

1. W. G. Sleight, "Memory and Formal Training", *Brit. Journ. of Psychol.* vol. 4, 1911, pp. 386-457.
2. C. Spearman, "Qualified and Unqualified 'Formal Training'", *Journ. of Exper. Pedag.* vol. 2, 1914.
3. E. L. Thorndike, *Educational Psychology*, vol. 2, 1921.
4. G. H. Thomson, *Instinct, Intelligence and Character*, 1924, chap. 14, The Transfer of Training.
5. A. I. Gates, "A Critique of Methods of Estimating and Measuring Transfer of Training", *Journ. of Educ. Psychol.* vol. 15, pp. 545-558.
6. P. B. Ballard, *The Changing School*, 1925, chap. 11, Formal Training.
7. E. L. Thorndike, "Mental Discipline in High School Studies", *Journ. of Educ. Psychol.* vol. 15, 1924, pp. 1-22, 83-98.
8. C. R. Broyler, E. L. Thorndike and E. Woodyard, "A Second Study of Mental Discipline in High School Studies", *ibid.* vol. 18, 1927, pp. 377-404.
9. C. H. Judd, *Psychology of Secondary Education*, 1927, chap. 19.
10. D. Starch, *Educational Psychology*, 1927.
11. G. P. Meredith, "Consciousness of Method as a Means of Transfer of Training", *Forum of Education*, vol. 5, No. 1, 1927 (adapted from portions of M.Ed. Thesis, Univ. of Leeds).
12. G. M. Whipple, "The Transfer of Training", *27th Year Book of the Nat. Soc. for the Study of Educ.*, Part 2, 1928.
13. P. Sandiford, *Educational Psychology*, 1928, chap. 14, Transfer of Training: Reciprocal Improvement in Learning.
14. H. Sorenson, "High School Subjects as Conditioners of College Success", *Journ. of Educ. Research*, April 1929, presents a case for the retention of Latin on account of its disciplinary value.
15. Report on Formal Training in Brit. Assoc. Report of Bristol Meeting 1930.
16. C. Fox, *The Mind and its Body*, 1931, chap. 10.

CHAPTER 13

THE MEASUREMENT OF SENSITIVITY

Materials.—(1) Colour Comparator, obtainable from Messrs. C. F. Palmer (London), Ltd. This motor-driven apparatus allows two interlocking colour discs cut along a radius to rotate at high speed, and enables the proportions of the two discs to be altered at will during rotation. (2) The following Hering standard colour discs, Red No. 2, Yellow No. 4, Green

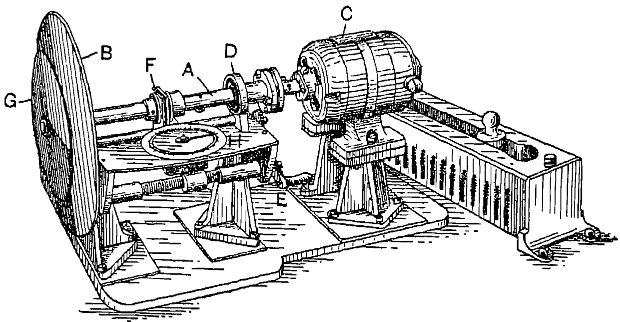

Fig. 21.

No. 7, Blue No. 10, together with sheets of thirty gray papers graded from black to white, *e.g.* those supplied by Zimmermann.

Section 1

THE MEASUREMENT OF ABSOLUTE THRESHOLD FOR COLOUR SATURATION

This measurement is affected by many factors which cannot here be discussed, but attention may be drawn to the following points:

(1) If, for example, the threshold for red is required, it is advisable to have the overlapping disc of a gray of the same brightness as the red. This should be determined by a previous

experiment as follows: Using one of the psychophysical methods described in Chapter 24, the proportion of black to white in the accompanying figures should be determined so that the resulting gray should match the red in brightness. This gray will correspond to one of the series of gray papers, and this gray should be used not only for the overlapping disc but also for the background.

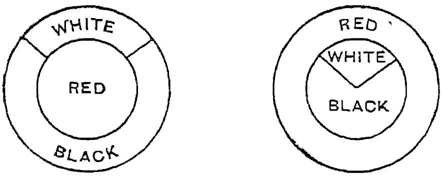

Fig. 22.

(2) The opening in the background should be just sufficient to allow the discs to rotate without touching the background. This seems to eliminate the contrast colour, green. It is not advisable for the red disc to have a smaller radius than the gray disc, as the contrast fringe of green surrounding the red disc would serve as a clue to S of the presence of red.

(3) S should be asked to judge whether there is a reddish colour present or only gray. He therefore knows that no other colour will appear in that particular test. This knowledge alone will radically affect the size of the threshold.

Section 2

THE MEASUREMENT OF THE DIFFERENTIAL THRESHOLD FOR COLOUR SATURATION

Procedure.—The coloured discs should be mounted as in the diagram. The outer part consists of two overlapping discs showing 180° Gray and 180° Blue of the same brightness as the gray. The inner part is equal in area to the outer, and the proportions of gray and red exposed can be varied. The apparatus is set in rotation, the screen is lifted and S judges whether the inner part contains more blue, less blue, or the same amount of blue as the outer.

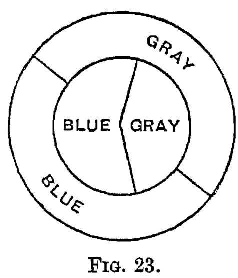

Fig. 23.

The Constant method described in Chapter 24 may be used.

Supplementary Experiment

The above experiment should be repeated (*a*) when the outer discs are adjusted so that 90° of Gray and 270° of Blue are exposed, and (*b*) so that 270° of Gray and 90° of Blue are exposed.

The differential threshold should be calculated for each case and compared with that obtained in the original experiment. The results should then be studied with reference to Weber's Law (cf. Chapter 24).

DISCUSSION

Some individuals are much more sensitive than others to a particular stimulus. It was said that Field-Marshal Earl Roberts could detect the presence of a cat hidden in a room, as it made him ill at ease. Be that as it may, there are several kinds of sensitivity whichever sense organ is considered. For example, colour sense or colour sensitivity can have many meanings. Of these, three may be mentioned here:

(1) Colour sense is sometimes taken to denote the capacity to see a minimal amount of a colour when mixed with black, white, gray or another colour.[1] The absolute threshold is a measure of this capacity. The artist mixing his colours or painting a picture and the chemist examining his precipitates or producing his dyes would be expected to find such a capacity a great asset. Certain railway employees should also possess it, as mere absence of colour blindness may not be a sufficient qualification in their case.

(2) Colour sense may also mean the capacity to distinguish between two shades of a particular colour.[2] The differential threshold is a measure of this capacity. It is recorded that employees in velvet factories on the Rhine could arrange in their correct order forty shades of black. There are also artists, milliners and workers in the cloth trade who exhibit marvellous powers of discrimination.

(3) In ordinary life one understands by colour sense the

[1] Cf. Ll. Wynn Jones, "Untersuchungen über die Reizschwelle für Farbensättigung bei Kindern", *Päd.-psychol. Arbeit.* vol. 2, 1911.

[2] Cf. W. Kobelt, "Untersuchungen über die Farbenunterschiedsempfindung bei Schulkindern", *Päd.-psychol. Arbeit.* vol. 5, 1914.

capacity to judge and evaluate colour combinations aesthetically. Civilised races differ from savages in their aesthetic judgments, children differ from adults. Even North European nations differ from each other in their colour schemes for dress or house decorations. As these three meanings of colour sensitivity may have little in common, it is necessary to study each by itself. To this end the significance of the term "threshold", which was introduced into psychology by Herbart,[1] must be learnt. "By the threshold of consciousness I mean those limits that an idea oversteps in passing from a state of complete inhibition to a state of reality." It is in one sense fitting that Herbart should have introduced such a term, for he himself was a living example of a threshold in so far as he stood on the dividing line between the metaphysical psychology which reigned before him and the scientific psychology which began to develop afterwards. It was Fechner, however, who developed this concept of the threshold in his epoch-making work on the psychophysical methods. By a threshold Fechner simply meant that point at which a stimulus or a stimulus-difference becomes just noticeable. For example, move a watch very slowly away from S's right ear and note the distance at which the ticking becomes inaudible. Now move it a yard further away and then approach S with it very slowly and note the distance at which the ticking becomes audible again. The mean of these two distances would be a measure of the absolute threshold for such a sound. Similarly, holding the watch at a specified distance smaller than this threshold, it would be possible to determine what distance must the watch be moved from this point in order that the sound could be definitely judged by S to be louder or fainter. This distance would be a measure of the differential threshold. Obviously each experiment ought to be repeated a large number of times.

THE ABSOLUTE THRESHOLD FOR COLOUR SATURATION

RELATION TO AGE.—Does the magnitude of the absolute threshold depend mainly on physiological conditions, probably inherited, in the structure of the optical system or on psycho-

[1] J. F. Herbart, *Werke*, 5, 292.

logical factors such as concentration of attention or the effect of practice? Wundt held that physiological factors are here dominant and that psychological factors were only of secondary importance. Yet I found that the absolute sensitivity for colour saturation is two to three times as great at fourteen years as it is at five. It must be acknowledged that the weight of evidence is against the possibility of physiological developments in the optical system going on during school years. If so, the increased sensitivity at fourteen, must be largely the result of psychological factors. But it is obviously difficult to produce any physiological evidence which would affect the issue.

Relation to Inheritance.—Amongst my subjects were a number of siblings and the results seemed to show that very often siblings not only possessed the same sensitivity to a particular colour but further that the sensitivities for red and for green run parallel, also those for blue and yellow, *e.g.* each of three brothers had less than average sensitivity for his age in the case of red and green, but greater than the average for blue and yellow.

Colour Weakness.—Such considerations enable us to give a definition of colour weakness which may have practical value. Children whose absolute threshold for a particular colour saturation lies appreciably below the average for their age are to be regarded as weak for that colour, although not in any sense subject to colour blindness. Such a test, when standardised, may have obvious vocational applications.

Relation to Introspection.—Various views have been expressed for and against admitting children as subjects in psychological experiments. Experimental pedagogy had been in existence for years before Wundt gave it his blessing, for he himself had used trained adults in all his researches. It all depends on the nature and difficulty of the introspection demanded. I came to the definite conclusion that in experiments like those described in Section 1, exact psychophysical methods may be used with the average child of five years, as he can be shown to be a reliable subject. The age of four can roughly be regarded as the limit.

Relation to Practice.—Practice does not seem to have

much effect on the magnitude of the absolute sensitivity for colour saturation.

THE DIFFERENTIAL THRESHOLD FOR COLOUR SATURATION

Sensitivity in this sense also increases with age, but on the whole the increase is not so marked as in the case of the absolute threshold. This is rather remarkable, as one would expect psychological factors to play a more dominant rôle in the case of the differential threshold. Kobelt nevertheless concludes that the increase in the differential sensitivity with age is due to the operation of psychological factors. When siblings were tested Kobelt found no evidence of family resemblances with respect to differential sensitivity. Further, practice was found to have an effect, but only when prolonged.

SENSITIVITY IN LIGHT-ADAPTATION

It is well known that some individuals take much longer than others to see objects when they pass from daylight into a darkened room. This difference has a vocational significance. The taxi-driver or the engine-driver may at one moment be in bright illumination and in semi-darkness at the next. Night flying aviators, in particular, should be quick adapters.

Bennett[1] studied the speed of adaptation from light to darkness of Standard VII. boys in Leeds. S's eyes were directed, without fixation, for three minutes on a large white screen, one metre distant, which was illuminated by a 580-candle-power lamp. This was deemed suitable, although the present writer had previously used a 1000-candle-power lamp for five minutes in his experiments with air pilots.[2] The lamp was screened so that its rays could not fall directly on S's eyes. The lamp was placed on a table near S's chair. Under the lamp, also one metre distant from the screen, was a small box con-

[1] E. F. Bennett, "Modern Developments to Safeguard the Vision of the School Child, with Some Experiments on Visual Adaptation", M.Ed. Thesis, Univ. of Leeds, 1927.

[2] Ll. Wynn Jones, "A Method of Measuring Nyctopsis", *Brit. Journ. of Psychol.* vol. 11, 1921.

taining an electric lamp, such as the Siemens 20 candle-power, gas-filled lamp (11·3 volts, 33 mm. bulb). In the front of the box, facing the screen, was a hole 1½ inches square, so that a patch of light could be thrown on the screen slightly below the level of S's eyes when a shutter was raised. The small lamp is connected with a suitable accumulator and an ammeter. At the end of three minutes the big lamp is switched off and a current of 0·6 ampere is sent through the small lamp, and S is instructed to say "now" when he can see the patch of light. To check the readings a cardboard with a hole in it was placed directly in front of the window of the box. The hole was either square, or triangular, or circular, and in each instance 0·5 square inch in area. The time taken by S to see the shape was noted. It is necessary to ascertain by preliminary experiments what is a suitable current for the particular lamp used. It is also assumed that an absolutely dark room is available for the experiment.

In a class of 31 boys, Bennett found that the quickest adapter saw the light after 1 min. 21 sec., while the slowest took 7 min. 25 sec.; the Quartiles being 2 min. 17 sec. and 5 min. 15 sec. and Median 3 min. 40 sec.

The corresponding figures for shapes correlated almost perfectly with those for light, the Quartiles being 3 min. 43 sec. and 5 min. 49 sec. and Median 4 min. 39 sec.

Bennett also found no correlation between the speed of adaptation and visual acuity as tested in daylight by the ordinary Snellen Test; a result which is in agreement with those obtained by Flugel,[1] and by the present writer.

QUESTIONS

1. Show how you would proceed to measure an absolute threshold and a differential threshold in the field of audition (cf. E. B. Titchener, *Experimental Psychology*, vol. 2, Part 1, 1905).
2. Using a bag of flour suitably weighted, show how you could measure S's ability to hit a mark distant 10, 20, 30, 40 and

[1] J. C. Flugel, "A Minor Study of Nyctopsis", *Brit. Journ. of Psychol.* vol. 11, 1921.

50 feet. Discuss the results with reference to Weber's Law (cf. Chapter 24).

3. Using golf balls and a jigger, mashie or mashie-niblick, measure S's ability to approach a flag distant 20, 40, 60, 80 and 100 yards. Discuss the results with reference to Weber's Law (cf. Chapter 24).
4. Measure S's visual acuity, for each eye, using Snellen's Test Types and employing Spearman's Method of "Right and Wrong Cases" with the intervals unequal. (This question is for the advanced student. Cf. *Brit Journ. of Psychol.* vol. 2, 1908, p. 227.)

References

1. E. B. Titchener, *Experimental Psychology*, vol. 2, Parts 1 and 2, 1905.
2. C. S. Myers, *A Textbook of Experimental Psychology*, 3rd ed. 1925.
3. G. M. Whipple, *Manual of Mental and Physical Tests*, Part 1, 1914.
4. M. Collins and J. Drever, *Experimental Psychology*, 1926.

CHAPTER 14

EXPERIMENTAL AESTHETICS AND ART JUDGMENT

Experiment 1: *Colour Preference*

MATERIALS.—A supply of coloured papers should be available.

METHOD.—*The Method of Paired Comparisons.*—Suppose the colours to be compared are spectral Red, Orange, Yellow, Green, Blue, Indigo and Violet. E prepares a set of squares of these colours, 4×4 cm. and a neutral gray cardboard 20×20 cm. pierced in the middle by two windows 3×3 cm. and 8 cm. apart. Each colour should be compared with every other. S sits at a table with closed eyes. E lays two colours on the table and places the neutral frame in position over them. At the word "Now" S views them for two seconds and then gives a decision as to which is the more pleasant. To ensure that all combinations are taken, the following chart is useful. Note that each colour comes twice in succession. Let it be at first on the right and then on the left in order to eliminate the space-error.

	I	II	III	IV	V	VI
II	1					
III	2	3				
IV	12	4	5			
V	13	14	6	7		
VI	19	15	16	8	9	
VII	20	21	17	18	10	11

FIG. 24.

Titchener's Paired Comparison Table (E. B. Titchener, *Experimental Psychology*, vol. 1, Part 1, p 93).

Let the Roman numerals I, II, III, IV, V, VI, VII denote the colours R, O, Y, G, B, I, V. Let the Arabic figures denote the number of the test, *e.g.* in square 1 appears the number of the colour preferred in comparison of II (left) and I (right).

If it is not desired that the same colour should appear consecutively, the following arrangement may be employed. Let Arabic figures now denote the colours. Arrange the combinations as in the following table:

12					
13	23				
14	24	34			
15	25	35	45		
16	26	36	46	56	
17	27	37	47	57	67

It is now easy to see that the following sequence will prove effective: 12, 34, 56, 13, 24, 35, 46, 57, 14, 25, 36, 47, 15, 26, 37, 16, 27, 45, 67, 23, 17. But in order to avoid the space-error it will be necessary to repeat the test with 21, 43, 65, etc.

DISCUSSION

It is clear that the colour sense involved in the colour preference experiment differs fundamentally from the two meanings given to the term in the last chapter. Sensitivity cannot now be measured by means of a machine. It is a question of taste or individual preference. Individuals differ widely and for various reasons in their preferences for colours. Not only so, but some are so sensitive to colours that they find no difficulty when asked to say which of two colours is more pleasant. Others, however, are so blunt or insensitive that the task appears meaningless to them.

Bullough[1] found that some individuals were more affected by the *objective aspect* of a colour, *e.g.* it was displeasing because foggy; others more by the *physiological aspect, e.g.* it was pleasing because soothing; others more by the *associative aspect, i.e.* the

[1] E. Bullough, "The 'Perceptive Problem' in the Aesthetic Appreciation of Single Colours", *Brit. Journ. of Psychol.* vol. 2, 1906-8.

colour was pleasing or displeasing because of the things of which it reminded them; still others more by the *character aspect, e.g.* a colour is pleasing because truthful or displeasing because aggressive. This, of course, must not be taken to mean that individuals can be divided into types. It is perhaps not necessary now to ascertain to what degree the behaviour of the subject in these experiments may be regarded as emotive, it need only be noted that the methods used for investigating emotive behaviour are here applicable. At this stage a brief account of such methods must suffice. They can be divided into two classes, methods of impression and methods of expression. In the former the subject gives an introspective report on his feelings when a stimulus such as a colour has been shown to him, or else the relative intensity of feeling is estimated when one stimulus is paired against another. In the latter[1] an attempt is made, usually by complicated apparatus, to note the bodily changes which have occurred, such as changes in breathing, in the rate and amplitude of the pulse, in blood pressure, in heart action, in the psycho-galvanic reaction, in reaction times, in the strength and amplitude of muscular contractions, in salivary secretions or in the discharge of glycogen from the liver into the blood stream. In many cases it is advisable to use the method of impression and the method of expression simultaneously. It will be convenient to deal with the former first.

METHODS OF IMPRESSION.—(1) *The Method of Choice.*—This is also known as the order of merit method. Objects such as advertisements or pictures are simultaneously exhibited and the subjects are asked to assign the value of 1 to the most pleasing, 2 to the next pleasing, and so on. It is obviously a rough method if only for the fact that so much is left to the whim of the subject, even though the objects are presented in a constant spatial arrangement.

(2) *Method of Serial Judgments.*—Each object is presented singly, after the subject has learnt a scale of absolute affective values such as: 1, very pleasant; 2, moderately pleasant; 3, just pleasant; 4, indifferent; 5, just unpleasant; 6, moderately un-

[1] Cf. P. M. Symonds, *Diagnosing Personality and Conduct*, 1931, chap. 11, Physiological Measures of the Emotions.

pleasant; 7, very unpleasant. As Myers[1] points out, this method serves as a valuable control over the method of paired comparison, for although one object may be preferred to another, yet absolutely the subject may be indifferent to it, or may even dislike it.

(3) *The Method of Paired Comparison.*—This in many ways is the most searching, and is the method employed in the last experiment. When a standardised scale of colours is in general use it will be possible to compare groups with respect to race, age, sex, culture or occupation. As that has not yet materialised it would take too much space even to summarise the results of various experimenters using different colours. Collins and Drever[2] refer to an investigation where men and women students put the saturated colours of the Bradley series in the following order of preference:

Women: Blue, Violet, Green, Yellow, Red, Purple.
Men: Blue, Violet, Purple, Red, Yellow, Green.

With young children yellows and reds would take a higher position.

Experiment 2: Art Judgment

MATERIALS.—The Meier-Seashore Art Judgment Test. Most persons complete the test within fifty minutes. The material consists of a book with 125 pairs of pictures, and may be obtained from the Bureau of Educational Research and Service, University of Iowa. The price is very reasonable, viz.: 1 copy for $1; 10 copies for $8·75; 35 copies for $28·5; 100 for $75. Each S should complete the test according to the instructions given with the test. Without acquainting the Ss with the results as given in the key, the present writer has found it instructive to ask each S to repeat the test after a week's interval with the added instruction to give a reason for his preferences. It will be found that the judgment has been altered in certain cases. Particular attention should be given to those cases and the reason for the change should if possible be ascer-

[1] C. S. Myers, *Textbook of Experimental Psychology*, 1911, Part 1, p. 307.
[2] M. Collins and J. Drever, *Experimental Psychology*, 1926, p. 203.

tained. By repeating the test in this way S may find it easier to analyse his judgments.

DISCUSSION

Three acts are usually held to distinguish the human race from the rest of creation; the search for truth, the recognition of moral laws and the aesthetic appreciation of beauty. Attention will here be confined to the last.

Even the primitive savage does not devote all his days to the struggle for existence or to activities which are merely utilitarian. Deep in his nature are cravings for beautiful objects. He will barter valuable commodities or face dire perils for a mere trinket which has captivated his aesthetic sense. In other words, he is influenced aesthetically by certain objects which correspond to the "objects of art" of his more civilised brother. Crude as his judgment of such objects may be, it is nevertheless definite and differs in important respects from his judgment in capturing his quarry or fighting his enemies.

In the laboratory also, if two tones differing in pitch are sounded in succession there is no difficulty in deciding by objective means whether S's judgment is correct or not. Similarly with regard to the more noetic judgments which are demanded, for example, in a test of intelligence or at a school examination.

If, however, two pictures are presented and one is asked to indicate a preference, how can it be ascertained whether the preference is "correct"? Who is to set the standard? In fact, is it at all certain that there is a standard?

If two paintings are recognised as masterpieces this difficulty will be easily apparent when they differ much in content. Thus a comparison of Hals's "Laughing Cavalier" with Turner's "Crossing the Brook" would appear futile in most aesthetic respects. But even if the two differ much less with regard to the kind of subject treated one may often experience the futility of demanding a preference. For example, a comparison of a landscape by Corot with one by Constable. It would appear that whatever art criteria be adopted, such a task is bound to prove difficult or even impossible. Two

different artists will differ in aims, outlook and technique, and even two pictures by the same artist cannot, for similar reasons, be compared with facility. Moreover, it does not in the least follow that if one expressed a preference for one painting over another on a given day that he would always do so; there is the influence of factors extraneous to the pictures themselves to be considered. Thus if one had to decide between a holiday on skis in the Erzgebirge and a trip to Madeira, each may be regarded *sui generis* as approaching perfection, but the decision actually made will probably depend on factors such as time, cost, possible companions, the desire for vigorous exercise or for rest and the like. Similarly in the comparison of paintings when one subject preferred the one which would go better with his furniture. In the face of such difficulties it is not surprising that experimental aesthetics has had to face formidable criticisms, and there is an unwarranted tendency to assume that the criticisms are fatal to experimental aesthetics, especially by authors who by temperament would not be expected to be very sympathetic. Thus William James wrote: "It strikes me that no good will ever come to Art, as such, from the analytic study of aesthetics—harm rather, if the abstractions could in any way be made the basis of practice".[1] But that is far from maintaining that the analytical study of aesthetics is futile, and even as it stands, the quotation is open to serious objection. No one has ever proved that analytical study hampers the work of the artist. The weight of evidence would seem to show that it may materially help him. Robert Bridges expressed this very clearly: "It (the natural impulse) may come to perfection only after long conscious toil and difficulty—and the sort of toil is different in the different arts. In all of them the Reason is a most active helpmate, but always the servant of the emotion. It is a mark of the consummate artist that the more he works on his production, the more he 'touches it up', the more 'spontaneous' it will appear."[2] In agreement with this would appear to be Goya: "Fancy abandoned by reason produces impossible monsters; united

[1] W. James, *Letters*, vol. 2, p. 86.
[2] R. Bridges, *The Necessity of Poetry*, p. 39.

with it, she is the mother of all the arts and origin of its wonders." Critics are apt to forget that the artist may be at one time a creator oblivious of his noetic reasoning and at another time a critical judge of his own production, and as a result continually improving it. Such considerations will enable us to make the necessary reservations before accepting the views of Croce[1] as to the nature of art. Croce denies—

(1) That art is a physical fact, *e.g.* as consisting of certain relations of lines such as the golden section.

(2) That art has to do with the useful or with pleasure and pain.

(3) That art is a moral act.

(4) That art has the character of conceptual knowledge.

In fairness to the critics it must be acknowledged that the task of postulating art criteria is formidable. In recent years much has been heard of the effects of "conditioning" and probably all would agree that aesthetic judgments might be greatly influenced—either for better or for worse—by conditioning. Hence the enormous importance of ensuring that the artistic stimuli presented to young children should be the most suitable. And such conditioning might well be with special reference to form or to content. Not only so, but mere habituation often suffices to modify the agreeableness of a presentation. Thus it has been said that the appreciation of quarter-tone music can be developed among Occidentals by repetition. Such considerations, however, are hardly important enough to preclude the possibility of universal standards. Clearly if a test of art talent is to be seriously considered there must be some discussion of the art criteria involved. Meier recommends "those tried and age-tested principles of art which have come down to us as balance, harmony and rhythm with the related ones of stability, symmetry, proportion and unity".[2] Meier also insists that the criterion of artistic rightness should be unimpeachable. The art world should afford its sanction.

[1] B. Croce, *Aesthetic*, trans. by Ainslie; also *The Essence of Aesthetic*, trans. by Ainslie.

[2] N. C. Meier, "Can Art Talent be Discovered by Test Devices?" Paper read at the 23rd Annual Meeting of the Western Arts Association, May 6, 1927, at Milwaukee, Wis.

The test device adopted in the Meier-Seashore Art Judgment Test is to require a comparison between test items of different value as judged by competent judges, and further the content has been kept so much under control that only one element is varied at a time. By this means the need of comparing two works of art is avoided, for, strictly speaking, only one of each pair is a work of art, the other violates in some particular one of those age-tested principles with which works of art should conform in all ages. Taste fluctuates, styles of painting change and theories of art multiply but the great picture survives. In the words of Delacroix: "The true beauties of art are eternal and will be accepted by all ages".

FURTHER EXAMPLES OF EXPERIMENTS.—Bulley[1] tested subjects from six to eighteen years of age with four pairs of pictures and asked them which picture in each pair they preferred. A marked decline in taste was noted as the small child grew older. Suggested reasons for the decline are given, but the conclusions are only put forward as provisional; and wisely so, as there were only four pictures. Gordon[2] used fifty coloured plates of Oriental rugs which were arranged by judges into an order of merit. Then the 1st, 3rd, etc., made up Series I., and the 2nd, 4th, etc., Series II. The pictures of a series were placed in chance order, and S had to arrange them in order of their beauty. He was also told not to think of theories of art but to judge them by the pleasure or displeasure they gave. There was a wide diversity of judgment, and this was more marked among a group which consisted of twenty experts. Nearly every rug was placed at the top and at the bottom by different judges. Some individuals agreed with the group judgment much better than others, and if they did so in one series they tended to do so in the other. Between one group and another the measure of agreement is large, and this increases with the size of the group. But this careful investigation by a competent psychologist raises queries which demonstrate the inherent difficulties of the problem. If each rug was a work of

[1] Margaret H. Bulley, *Art and Counterfeit*, 1925, p. 85.

[2] Kate Gordon, "A Study of Aesthetic Judgments", *Journ. of Exp. Psychol.* vol. 6, 1933.

art, why ask the subjects to rank them? For in such case it might be argued that experts would hardly be likely to agree, and that what is surprising is the large degree of agreement found between groups of individuals who were not experts, especially as a judgment in terms of pleasure is for many obvious reasons unreliable. And as regards the group order, attention must of course be paid to the composition of the group. Gordon tested 207 subjects. About one-half were undergraduate students of the Carnegie Institute of Technology. The majority of the rest were men and women with no academic connection, but probably above the average in artistic training.

Karwoski and Christensen[1] tested art appreciation in painting, architecture, industrial arts, abstract designs and colour, using three methods: (1) Comparing two examples, one good and one bad. The answer is covered by five specific reasons, only one being correct. (2) A single picture is judged as good or bad, again checking one out of five reasons. (3) Selecting the best of four examples of similar subjects.

As students of untrained groups overlapped into the highest group, the authors suggest that art taste is native or early acquired. Further, a correlation of about 0·28 between the tests and Thurstone's test of intelligence was obtained.

Another recent study is that of R. Alice Drought,[2] who accepts Hegel's definition of art as "the free and adequate embodiment of an idea, in a manner peculiarly adapted to the idea itself". The fundamental criteria chosen were unity, harmony, proportion and congruity which are tentatively defined. A test battery of nine sets of pictures, five to a set, were made by experts. Each set dealt with one specific subject and consisted of the perfect picture; one violating unity by creating two centres of interest; one violating harmony by omission of any suggestion of rhythm, or suggestion of relatedness of parts other than a feeling of oneness and spatial relationships; one violating proportion by lack of balance in the composition;

[1] T. F. Karwoski and E. O. Christensen, "A Test for Art Appreciation", *Journ. of Educ. Psychol.* vol. 17, 1926.

[2] R. Alice Drought, "A Survey of Studies in Experimental Aesthetics", *Journ. of Educ. Research*, vol. 20, No. 2, 1929.

one violating congruity by introducing superfluous features or features unsuited to the general character of the composition. The paired comparison method was used. Individual choices included all possibilities from choice of best picture to choice of the incongruous. In general, the order of preference was: perfect picture, that violating unity, that violating harmony, that violating proportion, the incongruous picture. Training in art causes an increase of sensitivity but the various violations were still found to be in the same order. It is interesting to note that the younger children, up to about eleven years, placed the incongruous picture first, but the adult and juvenile groups were less annoyed by a violation of unity than by a violation of harmony or proportion. This is largely corroborated by Jacob.[1] Meier-Seashore pictures were divided into three classes according as they violated balance, harmony or unity. It was found that secondary school boys, between the ages of eleven and seventeen inclusive, attending the City of Leeds School scored least marks on pictures violating unity; and up to the age of fourteen they scored most marks on pictures violating balance.

ART AND NOEGENESIS.—Spearman[2] has recently applied his fundamental principles to the analysis of artistic creations. His law of energy, which enounces that the quantity of mental output must be regarded as constant, bids the artist to eliminate everything irrelevant to his aim. One is thus reminded of Alberti's view: "I shall define beauty to be a harmony of all the parts in whatsoever subject it appears, fitted with such proportion and connection that nothing could be added, diminished or altered but for the worse".[3]

The law of energy also bids the art teacher select pictures which do not unduly tax the mental span of the scholars.

Further, Spearman's principle of retentivity which demands repetition must be supplemented by his principle of fatigue which requires variety. His fourth principle is that conation or striving intensifies cognition or knowing, and an attempt

[1] K. K. Jacob, "Art Judgment in School Children", M.A. Thesis, Univ. of Leeds, 1931.

[2] C. Spearman, *Creative Mind*, 1930.

[3] Cf. M. H. Bulley, *Art and Counterfeit*, 1925, p. 14.

is made to show how this principle functions in art. The fifth principle is that of "primordial potencies" which accounts for the widely different responses of different groups of individuals to works of art. Under this heading may be studied the differences due to Age, Sex and Race.

QUESTIONS

1. Give introspections of your judgments in specified examples of the Meier-Seashore Test.
2. Do you find it possible to classify the pictures in the Meier-Seashore Test according to principles which are violated?
3. Strict symmetry is avoided in art, yet a painting must be symmetrical. Explain.
4. What do you understand by the terms: golden section, the heaviness principle, empathy?
5. Discuss any psychological differences which may distinguish pictures appealing to children of about nine years from those which appeal to those of about thirteen.
6. "But the judgment of taste is not founded on concepts and is in no way a cognition, but only an aesthetic judgment" (Kant, *Analytic of Sublime*). Discuss the above statement.

References

1. E. Bullough, "On the Apparent Heaviness of Colours", *Brit. Journ. of Psychol.* vol. 2, 1906-8, p. 111.
2. E. Bullough, "The 'Perceptive Problem' in the Aesthetic Appreciation of Single Colours", *ibid.* p. 406.
3. E. Bullough, "The 'Perceptive Problem' in the Aesthetic Appreciation of Simple Colour-combinations", *ibid.* vol. 3, 1910, p. 406.
4. E. Bullough, "The Relation of Aesthetics to Psychology", *ibid.* vol. 10, 1919, p. 43.
5. C. W. Valentine, *The Experimental Psychology of Beauty*, 1919.
6. E. Bullough, "Recent Work in Experimental Aesthetics", *Brit. Journ. of Psychol.* vol. 12, 1921, p. 76.
7. C. K. Ogden, I. A. Richards and J. Wood, *The Foundations of Aesthetics*, 1922.
8. K. Gordon, "A Study of Aesthetic Judgments", *Journ. of Exp. Psychol.* vol. 6, 1923.
9. E. Newcomb, "A Study of the Appreciation of Beauty in School Children", *Forum of Educ.* vol. 2, 1924.

10. M. H. Bulley, *Art and Counterfeit*, 1925.
11. N. C. Meier, *Aesthetic Judgment as a Measure of Art Talent*, Univ. of Iowa Studies, vol. 1, No. 19, 1926.
12. N. C. Meier, "Special Artistic Talent", *Psychol. Bulletin*, vol. 25, 1928.
13. N. C. Meier, "A Measure of Art Talent", *Psychol. Mon.* vol. 39, 1928.
14. H. S. Langfeld, "The Rôle of Feeling and Emotion in Aesthetics" (being pp. 346-354 in *Feelings and Emotion*, Clark University Press, 1928).
15. C. Spearman, *Creative Mind*, 1930.
16. M. H. Bulley, *Have You Good Taste?*, 1933.

CHAPTER 15

PSYCHOLOGICAL TESTS IN MUSIC

Materials.—Seashore's Measures of Musical Talent comprising six 12-in. double records numbered A7536, A7537, A7538, A7539, A7540 and 53005-D obtainable together with a Manual of Instructions and Interpretations from the Columbia Gramophone Company, London.

Directions.—It is recommended, if time allows, that a laboratory period should be devoted to each of the above records, testing respectively the Sense of Pitch, Sense of Intensity, Sense of Time, Sense of Consonance, Tonal Memory and Sense of Rhythm. Failing that, E should take one or two of the records in detail and give a brief explanation of the others. To get reliable results it is important that the instructions which are given in the Manual should be rigidly followed. The following extracts will indicate the nature of the tests.

Sense of Pitch (100 judgments).—"You will hear two tones which differ in pitch. You are to judge whether the second is higher or lower than the first. If the second is higher, record H; if lower, record L."

Sense of Intensity (100 judgments).—"You will hear two tones which differ in loudness or strength. You are to judge whether the second is weaker or stronger than the first. If the second is stronger, record S; if the second is weaker, record W."

Sense of Time (100 judgments).—"You will hear three clicks marking off two intervals of time. If the second interval (that is, the time between the second and third clicks) is longer than the first interval, record L; if it is shorter, record S."

Sense of Consonance (50 judgments).—"You will hear two combinations of two tones each; one combination is better or worse than the other in consonance (harmony). A good combination is one in which the two tones are smooth, and blend, tending to fuse together into one. A bad combination is just the

opposite. If the second combination is better, record B; if worse, W."

Tonal Memory (50 judgments).—"In each trial you will hear a series of tones played twice. In the second playing, one note is changed. You are to record, by number, which one was changed."

Sense of Rhythm (50 judgments).—"You will hear in rapid succession two rhythmic patterns; the second is either the same as the first or different. Listen and record either S, for same, or D, for different." (The examiner will illustrate what is meant by rhythm.)

When a test has been concluded, each S should correct his mistakes, the correct answers being given orally by E from the key. It is, however, instructive, if time allows, to repeat a test with the same Ss on another day; in that case, it would be advisable not to mark the tests until the second has been given.

DISCUSSION

They are happy men whose natures sort with their vocations.
F. BACON, *Of Nature in Men*

It is now recognised that a good gramophone reproduces the artistry of a Caruso so accurately as to give enjoyment to discerning musicians. Further, thanks to the genius of Professor Seashore, it has materially aided psychological research in the field of music. The present writer applied the Seashore tests to 54 Standard VII. girls in Leeds with the following results:[1]

(1) The correlation coefficients between the first and second applications were:

Pitch = 0·75	Consonance = 0·39
Intensity = 0·44	Tonal Memory = 0·86
Time = 0·40	

These figures were obtained by changing the R-values of the

[1] Results distributed at a Meeting of Educational Section of Brit. Psych. Soc. in December 1923. At this time the Seashore Test for the Sense of Rhythm was not available.

Foot-Rule formula (cf. Chapter 23) into r-values by means of a table where

$$r = 2 \cos \frac{\pi}{3} (1 - R) - 1.$$

Although only approximate, they indicate the high reliability of the tests for the Sense of Pitch and for Tonal Memory.

(2) Fifty of these girls were also tested with Professor Spearman's Group Oral Tests of General Ability and the results correlated as follows:

Pitch = 0·40
Intensity = 0·34
Time = 0·29
Consonance = 0·01
Tonal Memory = 0·30.

Thus each test with the exception of the Consonance Test correlates appreciably with "g". These coefficients would be still higher when corrected for attenuation.

Seashore also found a relationship between intelligence and his tests, but attributes it to (*a*) that some children are so backward mentally that they are not able to do justice to their actual psychophysical capacity, and (*b*) that if a child is bright in music alone, this consideration would tend to improve his rating in general intelligence. In our case (*b*) does not apply as intelligence was measured by group tests but (*a*) probably plays some part, not only here but in countless researches where performances in intelligence tests are compared with those in other activities. With the exception of the Consonance Test, there is evidence then of positive correlation with intelligence, especially in the case of Pitch.

(3) Intercorrelations.—If Pitch, Intensity, Time, Consonance and Tonal Memory be denoted by P, I, T, C and M respectively, the first row in the following table gives the intercorrelations obtained in the case of the fifty-four girls already mentioned, using also the same formulae:

PI	PT	PC	PM	IT	IC	IM	TC	TM	CM
0·33	0·47	0·20	0·49	0·47	0·33	0·48	0·26	0·48	0·34
0·22	0·40	0·21	0·42	0·41	0·35	0·42	0·27	0·43	0·35

As already mentioned, fifty of these girls were given an intelligence test. It is now instructive to study the intercorrelations after eliminating the factor of intelligence by the use of a partial correlation formula.[1] The second row in the table gives the partial intercorrelations, *i.e.* assuming all the girls to have the same general intelligence.

The probable error[2] of the above coefficients is approximately 0·1. If there is a factor common to these tests other than "g", the application of the tetrad criterion might detect it (cf. Chapter 23). But even so, its specification would remain a question for research.

Seashore and Mount[3] also found appreciable correlation between sense of pitch and tonal memory. This is not surprising, as some of the changed notes in the tonal memory test only differed by a half-tone from the originals.

(4) As the Consonance Test was not found reliable it was therefore subjected to detailed analysis. There were available a Standard VII. class of 57 boys, another of 53 boys, another of 59 girls, 36 graduate men students and 63 graduate women students in Leeds. A rough estimate of the difficulty of each item in the test may be obtained by regarding a judgment *very easy* when it receives 90-100 per cent of the votes, *easy* if 80-90 per cent, *medium* if 70-80 per cent, *hard* if 60-70 per cent, and *indeterminate* if 50-60 per cent.

Occasionally there were interesting differences between the judgments of boys and girls, of adults and children, of males and females. For those cases it would, of course, be inadmissible to pool the results. Especially dangerous would it be to assume that the judgments of children in the Consonance Test should conform to that of adults. Nevertheless, some broad conclusions may be drawn. It was found that the test could be divided into five parts with about ten judgments in each, corresponding to the degrees of difficulties mentioned. And the value of the items, which appear in the class designated very easy, may only be

[1] See Chapter 23.

[2] *Ibid.*

[3] C. E. Seashore and G. H. Mount, "Correlation of Factors in Musical Talent and Training", Univ. of Iowa Studies in Psychology, No. 7, *Psychol. Mon.* vol. 25, No. 2, 1918.

negative, viz. that if wrong judgments are prevalent for these, then S is not gifted in this direction.

It was also found that, *for each group tested*, the voting went decidedly against the Seashore key in the case of four judgments:

(1) When the Octave follows the Fifth it receives 89 per cent of the votes, which is a majority for the key, but when it precedes the Fifth, only 41 per cent, which represents a majority against the key.

(2) Also 68 per cent vote for the Major Third in preference to the Octave. This again does not accord with the key. But as Valentine[1] found the Major Third heading the list of pleasant chords, our result is explained if we presume that most subjects judged in terms of liking and not in terms of fusion as demanded by the instructions.

(3) When the Minor Sixth follows the Minor Third it secures 77 per cent of the votes, a majority in favour of the key. When, however, it precedes the Minor Third the voting is 74 per cent against the key.

(4) Another notable case is that of the Minor Seventh and the Major Second. When the Seventh precedes the Second the voting is 85 per cent in favour of the Seventh in agreement with the key. When reversed, the voting is 75 per cent against the key.[2]

A curious result was also obtained with the Minor Third and the Diminished Fifth. When the Diminished Fifth comes second it was preferred by 55 per cent, and when it came first, by 60 per cent. Thus in both cases the majority disagree with the key. The majority, it is true, is not very convincing but it is consistent for each of the five groups tested. If we consult Helmholtz's table of the comparative roughness of different binary combinations, it will be seen that the Diminished Fifth has a roughness of 28, and the Minor Third 20. So that in any case the comparison would not be expected to be very easy.

[1] C. W. Valentine, "The Aesthetic Appreciation of Musical Intervals among School-children and Adults", *Brit. Journ. of Psychol.* vol. 6, Part 2, 1913, p. 196.

[2] Ll. Wynn Jones, "Experimental Studies in Consonance and Rhythm", *Proc. of VIIIth Internat. Cong. of Psychology*, Groningen, 1926.

But the verdict of our subjects seems to show that the proverb of the early harmonists, *Mi contra Fa diabolus est*, no longer applies to modern ears.

Professor Seashore himself is well aware of the difficulties involved in the Consonance Test: "Since so much depends upon understanding the definition of consonance, this test must be handled with caution—for children, because they may not understand, and with musically trained persons, because it is difficult for them to lay aside the known functional laws of harmony".[1] Thus the instructions demand a judgment on smoothness, blending and fusion and usually there is no means of knowing the relative stress placed on each of these three by the person tested. Further, "feelings of like or dislike" and "theory or feeling of musical value" may, and actually do, affect the judgments of some Ss although they are warned in the instructions not to pay attention to them. Even if we agree with Professor Seashore in defining "more precisely our conception of consonance by basing the decision upon a judgment rather than upon a feeling and to specify as the basis for this judgment the criteria which we find involved", yet the difficulty remains, especially with children, that a complex judgment involving at least three variables is required. The mere fact of their presence may cause confusion even if the three are almost identical. It is significant, however, that the median percentile rank for the Consonance Test in the class of girls was 76 while for each of the classes of boys it was only 49. This requires further investigation. It may be a sex difference, notwithstanding that sex differences which can be definitely established by exact statistical methods are somewhat rare. Or there may be a stronger tradition for aesthetic studies in girls' schools, so that girls generally are more interested in musical appreciation.[2]

(5) The Seashore Tests exhibit individual differences in a remarkable way. "We are dealing here not with differences of double, triple, or quadruple merely, but with differences of a

[1] C. E. Seashore, *The Psychology of Musical Talent*, 1919, p. 154.

[2] Cf. Board of Educ. Report on Differentiation of Curricula between the Sexes in Secondary Schools, 1923, p. 105.

ten-fold, fifty-fold, and one-hundred-fold magnitude in actual quantitative measurement".[1] It becomes then a serious matter if too much is expected in music of a pupil sadly lacking in some musical endowment. "Thus, a person who ranks 100 (in the Pitch Test) may hear a difference of one two-hundredth of a tone, whereas a person who ranks 1 cannot hear a difference of less than a half tone. The former is more than a hundred times as keen as the latter. The person who ranks 50, and therefore average, can hear a difference of 2·7 vibrations; *i.e.* five-hundredths of a tone."[2] It is well to note that the standard is A, 435 vibrations (*i.e.* double vibrations), which is known as the standard in international pitch, and that the differences involved in the successive columns of the Pitch Test are 30, 23, 17, 12, 8; $\frac{1}{2}$, 1, 2, 3 and 5 vibrations, one vibration being equivalent to one fifty-fourth of a tone.

An approximate guide, according to Seashore,[3] in the vocational guidance of a Standard VII. child as far as the Pitch Test alone is concerned is, if he can distinguish—

below 3 vibrations, he may become a musician;
3-8 vibrations, he should have a plain musical education;
9-17 vibrations, he should have a plain education in music only if special inclination for some kind of music;
only 18 vibrations and above, he should have nothing to do with music.

As a result of his numerous investigations Seashore recommends his six tests as measures of musical talent, for he maintains that musical talent is a complex of many fundamental and relatively independent factors, and that the recognition of difference in *kinds* of talent is the crying need of musical education today. Thus each one of the Seashore tests is held to measure some fundamental factor of musical talent, but no two of the tests measure the same factor.

As a contrast to the view of Seashore may be mentioned that

[1] C. E. Seashore, "A Survey of Musical Talent in the Public Schools", Univ. of Iowa Studies, vol. 1, 1920, p. 29.

[2] C. E. Seashore, *Manual of Instructions for Measures of Musical Talent*, p. 8.

[3] C. E. Seashore, *The Psychology of Musical Talent*, pp. 42, 66.

of Révész[1] who distinguishes between musical talent as exhibited by a composer or performer and musicality which enables its possessor to enjoy music. He holds that talent of necessity presupposes musicality, but that an outstanding musicality may be inborn in certain individuals with only a modest degree of talent. He maintains that musicality cannot be regarded as a complex of fundamental factors. Révész has devised six tests of musicality and he regards each of his tests to involve musicality to a greater or less extent, just as each intelligence test in an intelligence scale involves general intelligence. They are:

(1) *Melodic Rhythm.*—Bars were played on the piano, and the child reproduced the rhythm by hand-clapping.

(2) *Absolute Hearing.*—A note was played on the piano, and the child tried to reproduce it by pressing the correct key. Eight notes were used altogether.

(3) *Relative Hearing.*—The interval to be tested was given melodically on the piano, then another note was struck, and the child was directed to transpose the given interval. Eleven intervals were used.

(4) *Chord Analysis.*—Chords consisting of two or three notes were played on the piano, and the child was asked to sing the component notes.

(5) *The Perception and Vocal Repetition of Melodies.*—"Der Abendstern" (Schumann) was played once on the piano, then the first movement was repeated, and the child was asked to sing it. Four movements were taken altogether. This test is considered by Révész to be very closely related to musicality.

(6) *Instrumental Repetition of Melodies* (*playing by ear*).—A simple Hungarian song, well known to each child, was played on the piano with one finger. Then the child went to the piano and tried to play it. The number of mistakes was noted.

A new turn has recently been given to the testing of musical ability by the researches of Lowery,[2] who emphasises the importance of distinguishing between the interpretative and

[1] G. Révész, "Prüfung der Musikalität", *Zeits. f. Psychol.* 85, 1920.

[2] H. Lowery, "Psychological Tests of Musical Ability", M.Ed. Thesis, *Univ. of Leeds*, 1927.

technical aspects of musical performance. After a careful critical review of all previous investigations on musical ability, Lowery describes four new tests which he has devised, namely, the Cadence Test,[1] the Phrase Test,[2] the Accent Test and the Memory Test.[3] More recently Lowery[4] has devised two further tests, namely, a Quality Test, where S has to distinguish between the tone qualities of a piano, violin, clarinet, cornet and flute, and a Creative Test.

A characteristic of these tests is that they are all based upon actual musical material. Moreover, they provide excellent material for lessons on musical appreciation apart from their use for testing purposes.

QUESTIONS

1. What methods have been used, or could be used, in order to determine whether the Seashore Tests are measures of native capacities or of achievements? (cf. Stanton and Koerth's study mentioned in the references).
2. The following twelve tests were taken with a large group: the six Seashore Tests, and six tests of "intelligence", namely, Analogies, Completion, Instructions, Opposites, Inferences, and Classification. What results would you expect from a study of the ensuing tetrad differences? Give reasons for your answer. (Cf. Chapter 23.)

References

1. H. Rupp, "Über die Prüfung musikalischer Fähigkeiten", *Zeits. f. angew. Psychol.* 9, 1914.
2. C. E. Seashore, *The Psychology of Musical Talent,* 1919.
3. C. E. Seashore, "A Survey of Musical Talent in the Public Schools", Univ. of Iowa Studies: *Studies in Child Welfare,* vol. 1, No. 2, 1920.
4. G. Révész, "Prüfung der Musikalität", *Zeits. f. Psychol.* 85, 1920.
5. J. von Kries, *Wer ist musikalisch ?*, 1926.
6. H. Lowery, "Psychological Tests of Musical Ability", M.Ed. Thesis, 1927 (on file at the Univ. of Leeds).

[1] H. Lowery, "Cadence and Phrase Tests in Music", *Brit. Journ. of Psychol.* vol. 17, Part 2, 1926.

[2] *Ibid.*

[3] H. Lowery, "Musical Memory", *Brit. Journ. of Psychol.* vol. 19, Part 4, 1929.

[4] H. Lowery, "Estimation of Musical Capacity", *Mem. and Proc. Manchester Lit. and Philos. Soc.* vol. 76, 1932.

7. Hazel M. Stanton and Wilhelmine Koerth, "Musical Capacity Measures of Adults repeated after Music Education", Univ. of Iowa Studies, First Series, No. 189, 1930.
8. R. M. Drake, "Tests of Musical Talent", Ph.D. Thesis, 1931 (on file at the Univ. of London).
9. J. Mainwaring, "Experiments on the Analysis of Cognitive Processes involved in Musical Ability and in Musical Education", *Brit. Journ. of Educ. Psychol.* vol 1, Parts 2 and 3, 1931.
10. Univ. of Iowa Studies in Psychology. Up to date thirteen volumes have appeared at intervals since 1897. They are recommended for the student interested in researches on the psychology of music.
11. C. S. Myers, "The Psychology of Musical Appreciation", *Mem. and Proc. Manchester Lit. and Philos. Soc.* vol. 77, 1933.
12. H. M. Stanton and W. Koerth, "Musical Capacity Measures of Children repeated after Musical Training", Univ. of Iowa Studies, New Series, No. 259, 1933.
13. J. Mainwaring, "Kinaesthetic Factors in the Recall of Musical Experience", *Brit. Journ. of Psychol.* vol. 23, 1933.

CHAPTER 16

MANUAL DEXTERITIES: THE CONTROL OF TIME, STRESS AND RHYTHM

MATERIALS.—A Kymograph. A Marey Tambour with a straw lever at the end of which is a barograph pen. A tapping key with a spring powerful enough to give approximately the same "touch" as that of a piano key. A circular hole is made in the base to accommodate the closely fitting metallic box A whose top is covered by a rubber membrane. The rubber tube B passes into the metallic box through a hole in its side and has its other end attached to the tambour, so that the transmitted vibrations may be registered on the kymograph. The kymograph actually used had a circumference of 50 centimetres and its rate was such that a point on its circumference travelled 1·1 cm. in one second. But so many experiments can be arranged that it is impossible to suggest the best size and the best rate for all. In the experiments described below the tapping key was not used, as a device made for the writer by Messrs. Chappell & Co., London, was available. It was a model of the action section of a piano. As it was not found practicable to measure the force of the blow of the hammer, it was nevertheless found, after preliminary experiments, that the apparatus could profitably be employed by boring a hole in the base and using a metal box as indicated above. But obviously the simple key would serve equally well.

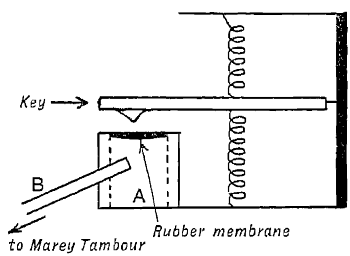

FIG. 25.

DIRECTIONS.—(1) *Control of Time Test.*—"Tap on the key with the first finger in time with the metronome. After eleven taps I shall stop the metronome but you must continue to tap

at the same rate as before, just as if the metronome was beating time. Keep on doing this until I tell you to stop. Do not tap too heavily."

This is done once by S for practice and is not registered on the drum. The trial is then repeated and registered. S should be told not to tap heavy enough to cause the lever to attain its maximum amplitude, for this, whilst not affecting the present test, would obviously vitiate the tests mentioned below. S is also requested to lift his finger from contact with the key after each tap. S actually taps twenty-one times, eleven with the metronome and ten without.

(2) *Control of Intensity Test.*—"Tap on the key with the first finger twelve times, each tap to be exactly alike as far as possible, always tapping with the same pressure or force and touching the key always at the same place."

This is done once for practice and is not registered. The trial is then repeated and registered. S is told that the first two taps serve to give him an idea of the force necessary to comply with the instructions and that only the last ten taps will be measured.

The following additional tests may be taken by students specially interested:

(3) *Control of Time and Intensity Test.*—Similar to Test 2, but with the added injunction that the tapping should be as regular in time as possible.

(4) *Trochee Rhythm Test.*—"Tap 'heavy, light' on the key with the first finger until told to stop. Keep regular time, that is, do not go faster or slower, also keep the pressure regular, *i.e.* every heavy tap should have the same force and every light tap the same (smaller) force. Be careful not to make the heavy tap any heavier than in the previous tests."

Ten pairs of "heavy, light" taps to be done for practice without registration and then ten pairs registered.

(5) *Dactyl Rhythm Test.*—A similar test to Test 4 with the pattern "heavy, light, light".

(6) *Five-finger Test.*—"Tap the same key with thumb, 1st, 2nd, 3rd and 4th fingers in succession, each tap to have the same force. Repeat this five times for practice without registering, then the test is done again with registration."

If there is a tendency to tap too lightly or too heavily with any finger, it will be shown on the record.

(7) *Minimum Pressure Test.*—It is recommended that the straw lever in this test should be at least eight inches long in order to magnify the tracing exhibiting the movements made.

"Tap on the key with your first finger so as to move it as little as possible. Keep doing this until I tell you to stop. Be sure that you move the key, but make the movement as small as you possibly can every time."

This is done twelve times for practice without registration and then twelve times with registration, the last ten taps being measured, *i.e.* the heights through which the lever moves each time is measured.

DISCUSSION

> How sour sweet music is
> When time is broke and no proportion kept.
>
> SHAKESPEARE, *Richard II.* V. 5

The apparatus described in this chapter enables the student to carry out diverse tests on the control of time, control of stress, control of rhythm, a comparison of the sensitivity of the different fingers, etc. The instructions can be carried out by children in Standard VI. and the performances of adults and of skilled pianists may be obtained for comparison.

Tapping with the finger is an activity which can be tested with regard to several aspects:

(1) The speed of tapping has received very much attention from psychologists. It is comparatively easy to measure the speed which it was thought would have many practical uses. Ream[1] mentions eighteen of the principal uses to which the test has been put.

(2) The regularity of tapping with respect to time has received far less attention. Over twenty years ago Seashore[2] at the

[1] M. J. Ream, "The Tapping Test: a Measure of Motility", *Psychol. Mon.*, 1922. Cf. also S. Nurullah, "A Study of Rhythm and Psychological Methods of developing Regularity of Time and Stress in Movement", M.Ed. Thesis, Univ. of Leeds, 1927 (on file); G. M. Whipple, *loc. cit.* Part 1, pp. 130-147.

[2] C. E. Seashore and G. H. Mount, "Correlation of Factors in Musical Talent and Training", *Psychol. Mon.*, 1918.

University of Iowa had devised apparatus for measuring the "time of free action" and the "time of regulated action". The regularity of the former was measured in terms of the average error in the free tapping of time at the rate of one tap per second; the regularity of the latter was measured in terms of the average error in keeping time with a click made at intervals of one second. By means of an ingenious adaptation of Seashore's phonograph chronograph Prof. R. H. Seashore[1] has recently developed a motor rhythm test by means of which a subject's temporal precision in tapping to a prescribed rhythm may be measured.

(3) The regularity of tapping with respect to intensity or stress is another important aspect which has received still less attention from psychologists.

(4) Finally may be mentioned the sensitivity in the sense of the delicacy in moving a key as little as possible.

It seems to the writer that the second, third and fourth measurements in their practical applications may far outweigh in their importance the first, the measurement of speed. A brief summary of tests taken by the present writer during the session 1922-23 will therefore be given:

The subjects were 48 Standard VII. boys from the Blenheim Elementary School, Leeds, and the writer is indebted to Mr. Whatmoor, the Headmaster at that time, for allowing the boys to visit the laboratory. In Test 1, the Control of Time Test, the range of the boys' variability[2] varied from 1·1 per cent to 10 per cent with

Upper quartile . .	3·2 per cent
Median. . . .	3·7 ,,
Lower quartile . .	4·5 ,,

In Test 2, the Control of Intensity Test, the range of variability extended from 2 per cent to 51 per cent with

Upper quartile . .	6·0 per cent
Median. . . .	10·5 ,,
Lower quartile . .	18·0 ,,

[1] R. H. Seashore, "Studies in Motor Rhythm", *Psychol. Mon.*, 1927.

[2] Cf. Chapter 23.

JUDGING TIME AND MARKING TIME

Moreover, these boys had taken the Seashore Test of the Sense of Time which requires the judging of time, whereas the marking of time is required in Test 1. The question therefore arises: Do those who excel in one of these tests necessarily or usually excel in the other? The percentile rank[1] of each boy in the Seashore Test for the Sense of Time was calculated and also the variability of each boy in the Control of Time Test (tapping). Thus two series of values were obtained, one for judging time and one for marking time. They yielded a product-moment[2] coefficient of correlation of 0·32 (probable error being 0·09). Thus the two abilities do not correlate very closely and this result might possibly have been predicted, for do we not sometimes meet individuals who excel in pointing out the trivial errors of others in some motor tasks although their own errors at the same task could by no means be regarded as trivial? Or a poet may extol the grace of the harvester by the phrase "rhythmic as the swish of the swinging scythe", although unable to wield a scythe himself.

Of these 48 boys, it was found that 14 were above the average of the class in judging time and below the average in marking time. But only 7 were above the average in marking and below the average in judging time. This suggests that if a person *can* mark time well, we have at least some hope of his judging time well; but if a person judges time well, we are not justified in having the same degree of assurance that he will mark time well.[3]

REGULARITY IN TIME AND REGULARITY IN INTENSITY

When the results of Test 1 and Test 2 were compared[4] it was found that there was no correlation between the ability to tap regularly in time and the ability to tap regularly in intensity. Moreover, no correlation was found between these abilities and "general intelligence". Whether these provisional conclusions

[1] Cf. Chapter 23. [2] *Ibid.*
[3] Cf. Ll. Wynn Jones, "Tests of Musical Ability", *Child Life*, June 1926, and "Experimental Studies in Consonance and Rhythm", *Proc. of VIIIth International Congress of Psychology*, Groningen, 1926. [4] *Ibid.*

would hold after prolonged practice remains, of course, undetermined.

METHODS OF TRAINING REGULARITY IN TIME

(1) VISUAL METHOD.—A horizontal line is drawn on the kymograph paper and a series of marks are made at regular distances which represent one second when the drum is moving at the speed used in the experiments.

Ss are instructed: "Note the marks on the line. In one second the drum moves so that the pen traces a line from one mark to the next. In tapping, try to get the pen to pass through each successive mark in its upward movement."

Ss in this group are allowed 30 taps daily for five days.

(2) AUDITORY METHOD.—Ss are instructed: "You will hear a metronome ticking seconds. Be careful to tap as nearly with the ticking of the metronome as possible."

Ss in this group are allowed 30 taps daily for five days.

(3) COUNTING METHOD.—Piano teachers often recommend the system of playing to the counting, so in this method Ss were advised to tap to their counting and were instructed thus: "You are to tap at regular intervals of a second. It will be helpful to you if you try to find out which counting will suit you best. It may be 1, 2, 3 or 1, 2 every second, or any such sequence. Having determined that, always tap to your counting and do not count to your tapping."

Ss in this group are allowed 30 taps daily for five days.

METHODS OF TRAINING REGULARITY IN INTENSITY

(1) VISUAL METHOD.—Two parallel horizontal lines are drawn on the drum, 3 cm. apart. The pen, when at rest, was on a level with the lower line. Ss are instructed: "You are required to tap with such force that the pen just reaches the upper line at each tap. You can do that only if you use a constant force. Pay no heed to regularity of time."

Ss in this group are allowed 30 taps daily for five days. As the drum was not screened S can see the result of his tapping and regulate his touch accordingly.

(2) Oral Instruction Method.—The two parallel lines were drawn as in the former method. But instead of having the drum visible it was screened. Instead of visual guidance, another kind of guidance was given, namely, verbal instructions. Ss were instructed:[1] "You are to use the same force every time you tap. When you use too great a force, I shall call out 'Big' and you will have to decrease it the next tap. If you use too small a force I shall call out 'Small' and you will have to increase it the next tap. If you use the right amount of force I shall remain silent which means that you are to use the same force again. Pay no heed to regularity of time."

Ss in this group are allowed 30 taps daily for five days.

(3) Self-rivalry Method.—This method can also be adapted for training regularity in time. Each S is shown his record after each practice of 30 taps and his errors are pointed out with a view to his improving each day on his previous performance.

Ss in this group are allowed 30 taps daily for five days.

All these methods, except the last, were used by Nurullah, who had four subjects in each of his training groups. When the performances of each group were compared with that of a control group, there was no definite evidence of the value of any of the methods after a fortnight's practice, the Ss being then retested with the original tests. It is true that some had improved but others had actually deteriorated. This is not so surprising in view of R. H. Seashore finding only a slight improvability in motor rhythm. And as regards improvability in the sensory rhythm test of C. E. Seashore, Miss Klaver[2] found that two months of intensive training of all sorts—marching, clapping, beating time, listening, etc., failed to show a significant rise in the average of the group. It should again be emphasised that in R. H. Seashore's motor rhythm test, S has to *tap*, whilst in C. E. Seashore's sensory rhythm test, S has to *judge*.

A Comparison of Motor and Noetic Variability.—In comparing the variability of an individual in judging rhythm with his variability in marking rhythm, one is inclined to con-

[1] Nurullah employed this method in 1926 (*op. cit.*) and could not therefore profit from a somewhat similar device used by Thorndike (cf. "The Law of Effect", *Amer. Journ. of Psychol.* vol. 39, 1927).

[2] R. H. Seashore, *op. cit.* p. 180.

clude that he is more variable with respect to his fellows in the latter than in the former. The general factor "g" tends to keep the former within certain limits, but considerable fluctuations are found in motor performances even in the case of persons possessing an innate motor ability which has been practised over a prolonged period. In other words: Man is not a machine. Common observation would seem to show that similar results hold for more complex motor activities, *e.g.* the long handicap golfer may experience both the thrill of breaking eighty and the annoyance at totalling a hundred strokes. But the formulation of the laws of motor variability is a task which has hardly yet begun and will require prolonged research into the nature of the specific factors which function in particular motor activities.

THE NATURE OF RHYTHM.—Originally "rhythm" seems to have denoted movement, probably of the sea or the river. Later on it included the concept of gracefulness, and was applied sometimes to things without movement, *e.g.* garments. It was also used by the Greeks analogically to denote moral or aesthetic characteristics, although a classical scholar [1] finds it difficult to distinguish between their use of the words "symmetry" and "rhythm". Passing on to modern times, two definitions of rhythm may be given. According to Sonnenschein: [2] "Rhythm is that property of a sequence of events in time which produces on the mind of the observer the impression of proportion between the durations of the several events or groups of events of which the sequence is composed." It will be noticed that "stress" is not explicitly mentioned in his definition, but he agrees that the ictus or stress or accent which is indicated by the bar-line, may form part of "that property of a sequence of events in time" (musical notes in this case).

Not very different is the definition of R. H. Seashore: [3] "Motor rhythm, as distinguished from other forms of orderly action, may be defined as a progression in action by balanced deviations in time, intensity, or quality from the simple periodicity of any regular recurrent action." However, in the R. H. Seashore motor rhythm test, which is an adaptation of

[1] E. A. Sonnenschein, *What is Rhythm?* 1925, p. 15.
[2] *Op. cit.* p. 16.
[3] R. H. Seashore, *op. cit.* p. 142.

the C. E. Seashore phonograph chronograph, the factors of intensity and quality are constant and the criterion of ability is temporal precision.

Objections to the use of tests of motor rhythm are sometimes based on the fact that no musical performer adheres to the time-values indicated by the score, but instead adds greatly to the expressiveness of the music by means of small variations of the tempo, in other words, by phrasing. Especially so in the case of an organist where phrasing is produced without the aid of variations in loudness. The reply to such criticism is that in order to produce the phrasing, the performer must obviously be able to control the time accurately enough to enable his listeners to appreciate his phrasing.

Time-values which differ from one another up to a certain limit may yet be regarded by the listener as equivalent. This is markedly the case in verse. Thus the following line of Tennyson when recorded on the kymograph by Sonnenschein[1] showed the following ratios of duration:

The long light shakes a-cross the lakes
12 : 31 : 27 : 45 : 7 : 34 : 9 : 55

These figures do not exhibit anything like a constant ratio, whether of foot to foot or of rise to fall within the foot, yet a certain impression of proportion between the durations is produced on the mind of the listener.

Even expert players in music by no means produce the theoretical values for the temporal relations indicated by the score. Yet their listeners are entirely unaware of the fluctuations. Thus Morton,[2] using an electrical tuning-fork making 128 vibrations per second, measured the degree of regularity attained by players on the piano. For the most expert player the average interval (over 22 notes) was 28·7 vibrations, the least 26·7, the greatest 30·0, and the standard deviation 0·96. At the other extreme was a player with an average of 32·7,

[1] *Op. cit.* p. 32.

[2] W. B. Morton, "Some Measurements of the Accuracy of the Time-intervals in playing a Keyed Instrument", *Brit. Journ. of Psychol.* vol. 10, 1920.

varying from 22·9 to 46·6, with a standard deviation of 5·56 vibrations.

In agreement are the results of Stetson and Tuthill [1] where the records of six musicians showed time divisions for the dotted eighth-sixteenth unit-group which varied all the way from the theoretical ratio 1 : 3 through 1 : 2 to 2 : 3. It is true that the type of the unit-group at a given tempo for a given sitting is fairly constant, but the type varied widely not only from subject to subject and from tempo to tempo but also from sitting to sitting at the same tempo with the same subject. Moreover, the unaccented note was decidedly longer than the theoretical value at all tempos and with all subjects.

Practical Ability is a term which has been used rather loosely by different writers. Sometimes it refers to innate aptitude or capacity, sometimes to acquired skill. Moreover, tests which involve bodily movements differ widely with reference to the relative importance of "intelligence" and manual skill for successful performance. The following rough classification may serve as illustration:

(1) Eductive Tests: (*a*) those making considerable demands on "g" but not on manual skill, such as Healy's Picture Completion Test,[2] some of McFarlane's [3] tests which were labelled tests of "practical ability", also some of Vickers and Hoskin's [4] which were labelled tests of "practical abilities".

(*b*) Those demanding "g" and also some manual skill such as Kelley's Test of Constructive Ability.[5]

(*c*) Those demanding "g" and also a "mechanical factor" but no manual skill (cf. Cox's Tests in the next chapter).

(*d*) Those demanding "g", one or more mechanical factors and some manual skill (cf. Stenquist's Tests in the next chapter).

[1] R. H. Stetson and T. E. Tuthill, "Measurements of Rhythmic Unit-groups", *Psychol. Mon.* vol. 32, 1923.

[2] W. Healy, "Pictorial Completion Test (2)", *Journ. of Applied Psychol.* vol. 5, 1921.

[3] M. McFarlane, "A Study of Practical Ability", Mon. Suppl. 8, *Brit. Journ. of Psychol.* 1924.

[4] W. Vickers and V. H. Hoskin, "Results from some New Tests of Practical Abilities", *Brit. Journ. of Psychol.* vol. 20, 1929.

[5] T. L. Kelley, "A Constructive Ability Test", *Journ. of Educ. Psychol.* vol. 7, 1916.

(2) Motility or Dexterity Tests depending more on specific factors and demanding strength, speed, delicacy of touch or steadiness, such as those of Perrin,[1] R. H. Seashore,[2] Earle[3] and others.

Extensive testing of various forms of motor ability has tended to disprove three views which were sometimes held. There is no general motor ability, neither can motor ability be considered dependent on a number of simple variables like speed of reaction, nor else on a relatively small number of basic motor capacities. On the contrary, motor abilities are highly specific when the influence of "g" has been eliminated.

QUESTIONS

1. What distinctions have been drawn between "skill" and "dexterity" when applied to describe the behaviour of an individual in a specific task?
2. What theories have been advanced to account for the observed individual differences in motor skills?

References

(additional to those mentioned in the text)

1. W. R. Miles, "A Pursuit Pendulum", *Psychol. Rev.* vol. 27, 1920.
2. C. F. Hansen, "Serial Action as a Basic Measure of Motor Capacity", *Psychol. Mon.* vol. 31, 1922.
3. B. Muscio, "Motor Capacity with special reference to Vocational Guidance", *Brit. Journ. of Psychol.* vol. 13, 1922.
4. G. Ehinger, "Recherches sur le développement de l'habilité manuelle par la pratique d'un métier manuel", *Arch. d. Psychol.* vol. 20, 1927.
5. W. R. Miles, "A Rotary Motility Test", *Journ. of Gen. Psychol.* vol. 1, 1928.
6. F. Akroyd, "Some Tests and Correlations of Hand-eye Co-ordination", *Forum of Education*, vol. 6, 1928.
7. H. Antipoff, "L'Évolution et la variabilité des fonctions psychomotrices", *Arch. d. Psychol.* vol. 21, 1928.

[1] F. A. C. Perrin, "An Experimental Study of Motor Abilities", *Journ. of Exp. Psychol.* vol. 4, 1921.

[2] R. H. Seashore, "Individual Differences in Motor Skills", *Journ. of Gen. Psychol.* vol. 3, 1930.

[3] F. M. Earle, assisted by F. Gaw and others, *The Measurements of Manual Dexterities*, Rep. 4. Nat. Inst. of Indust. Psychol., 1930.

8. A. Gemelli, "Recherches sur la nature de l'habilité manuelle", *Journ. d. Psychol.* vol. 26, 1929.
9. A. E. Weiss Long and T. H. Pear, *A Classification of Vocational Tests of Dexterity*, Rep. No. 64 of Indust. Health Research Board of Med. Res. Council, 1932.
10. J. N. Langdon, "An Experimental Study of Certain Forms of Manual Dexterity", Ph.D. Thesis (on file), Univ. of London, 1932. Cf. also Rep. No. 66 of Indust. Health Research Board, 1932.
11. E. M. Henshaw, P. Holman and J. N. Langdon, "Manual Dexterity, Effects of Training", Rep. No. 67, *ibid.* 1933.

CHAPTER 17

MECHANICAL APTITUDE

MATERIALS.—(1) The Stenquist Picture Tests of Mechanical Aptitude. Test I. contains pictures of mechanical objects with questions about relationships. Test II. is somewhat similar but more difficult and involves verbal responses. The tests may be obtained from the World Book Co., Yonkers-on-Hudson, New York, in packages of 25 with Key and Class Record. Test I., $1·50; Test II., $1·50; Manual of Directions, 15 cents; Specimen Set, 30 cents.

(2) The Cox Tests of Mechanical Aptitude. (*a*) Mechanical Models. (*b*) Mechanical Completion. (*c*) Mechanical Explanation. (*d*) Mechanical Diagrams as described in *Mechanical Aptitude*, by J. W. Cox (7s. 6d. Methuen & Co., 1928).

(3) The Stenquist Assembly Test of Mechanical Ability as revised by the National Institute of Industrial Psychology. Directions are given in *Tests of Mechanical Ability*, by F. M. Earle and A. Macrae and others (3s. 6d. National Institute of Industrial Psychology, Aldwych House, London, W.C.2, 1929).

Each of the above tests should be qualitatively discussed in class with a view to clarifying the concept of "mechanical aptitude". Definite conclusions as to the scope of the various tests or quantitative data as to their relative efficiency cannot be obtained in the time available on account of the complexity of the problem. Thus Cox has later found evidence that the "mechanical factor" present in his tests also plays a part in suitable tests of the mechanical assembly type. In addition, evidence has been obtained for a "routine manual factor" associated with manual tests, and this factor is distinct from the mechanical factor.[1]

[1] Cf. Rep. of Brit. Assoc. 1933, *Rep. of Committee on Factors involved in Mechanical Ability*, p. 306.

DISCUSSION

> After the machine is finished, and the steam turned on, the next is to drive it; and experience and an exquisite sympathy must teach him where a weight should be applied or a nut loosened.
>
> R. L. STEVENSON
>
> *Records of a Family of Engineers*, p. 212

The pioneer work on mechanical aptitude was carried out at Teachers' College, Columbia University, where Stenquist,[1] following a suggestion of Thorndike, experimented with the well-known Assembly Tests and Picture Tests which have been recommended in this chapter.

It is common knowledge that mechanical ability or "machine sense" is possessed by some persons to a far greater degree than by others. But when attempts are made to assess this ability by psychological tests, the difficulty of the task soon becomes apparent. It is obvious that the age of the subject must be considered, also his degree of practice and the nature of his environment. It is also not surprising to find that the subject's general mental ability will favourably affect his performance to an extent dependent on the nature of the test. But even if the subjects are of the same age and have had equal practice and a similar environment and possess the same measure of "g", it is still found that there are other factors which affect their performances with mechanical appliances in real life. Three of these factors deserve special consideration.

(1) MORAL CHARACTERISTICS.—There are certain moral characteristics which, although of universal importance, require special emphasis when dealing with applicants for posts involving complicated mechanisms. There must be a steadiness and persistence in the care of the mechanism, a patience in dealing with mechanical troubles, and truth, honesty and conscientiousness in all dealings with regard to it. The importance of such moral qualities was clearly seen in the results obtained by Greenwood[2] in his analysis of the qualities necessary for a chemist. A questionnaire was sent to fifty-two expert chemists,

[1] J. L. Stenquist, *Measurements of Mechanical Ability*, 1923.

[2] N. Greenwood, "A Vocational Study of Chemistry with its Applications to Education", M.Ed. Thesis, Univ. of Leeds, 1924.

namely, teachers at universities, technical colleges and secondary schools, city analysts, and industrial chemists engaged in gas manufacture, colour chemistry, textiles, organic chemistry, etc. They attached much more importance to these moral qualities than they did to native intelligence, experimental ability, memory, or such qualities as a steady hand or a keen eye.[1] The following illustrations are taken from Greenwood's account.

Case A.—A research chemist of some years' experience was experimenting with gases when his apparatus commenced to leak. He should have dismantled his apparatus and commenced a fresh "run". Instead he estimated the air leaking in as 3 c.c. Obviously he was no chemist.

Case B.—Owing to fear of trouble a chemist did not report a faulty working of a certain unit of a plant. He would have had the fireman "on his chest" he gave as his reason.[2]

At the risk of emphasising the obvious, another illustration from another field may be given: Assuming that a friend wishes to engage a chauffeur and that the following table gives a brief psychographic account of three chauffeurs, X, Y and Z, made by a psychologist who is acquainted with them.

	Moral qualities: Steadiness, Honesty	Driving Ability	Mechanical Aptitude	Genial Companion	General Intelligence	Appearance
X	C +	C	C	A	C	C
Y	A	B +	B	C	B	A
Z	D	A	A	B +	B +	A

If provisionally all the qualities in the table are weighted the same, and if it be agreed to award 5, 4, 3, 2, 1 marks for A, B, C, D and E respectively, and to add a third of a mark for a plus and subtract a third of a mark for a minus, it will appear as if Z might be recommended, but his would-be employer may rightly object to the scheme and reply that he does not particularly desire a boon companion or even high intelligence as long as he gets a man who is honest and can drive

[1] N. Greenwood, "A Vocational Study of Chemistry with its Applications to Education", M.Ed. Thesis, Univ. of Leeds, 1924, p. 69.

[2] *Ibid.* p. 36.

and look after his car, and he therefore decides to choose Y. Actually in this particular instance a judgment from general impression is very easily made. Y would undoubtedly be the best choice, and even X would be preferable to Z, which again goes to confirm the heavy weighting which should be attached to moral characteristics.

Lack of these moral qualities also causes the partial failure of certain psychological tests when applied in industry. A performance test may fail to correlate as highly as would be expected with the ratings of supervisors for the simple reason that some of the employees are careless in their work although under the stimulus of the test they show no lack of manual dexterity or whatever factor is operative in the test.

(2) MANUAL DEXTERITY.—This is another factor to be considered. Some machines, for example, a typewriter, demand a good deal of dexterity if little mechanical aptitude. Indeed, the predictive value of tests of manual dexterity may be such as to warrant their inclusion in the testing programme when testing applicants for various manual trades.[1]

(3) MECHANICAL APTITUDE.—Finally there is mechanical aptitude as defined by Cox, namely, a group factor running through those operations in which the subject is called upon to deal mentally with mechanical movements. It seems to be an innate aptitude which enables its possessor to think and act effectively in processes where the correlates and relations are spatial in character. "Such thinking involves not merely the apprehending of certain spacially arranged items,[2] as in a design, nor solely in the eduction of their space relations whereby the design is known to have a definite shape, or pattern: superimposed on this definite arrangement of parts is movement—and as the items of the mechanism move their pattern continually changes."[3] In this sense it is a factor additional to "g" but operating always in conjunction with "g" and distinct from manual dexterity. Cox points out that

[1] Cf. *Tests of Mechanical Ability*, Report 3 of the National Institute of Industrial Psychology, p. 39.

[2] Cf. W. Blumenfeld, "Untersuchungen über die Formvisualität", *Zeitsch. f. Psychol.* 91, pp. 1-82, 236-292, 1922.

[3] J. W. Cox, *op. cit.* p. 161.

mechanical aptitude, thus defined, has little scope in the activities of the usual school subjects, and it is therefore difficult to judge whether a boy leaving school is really fitted for an engineering occupation which demands high mechanical aptitude. Hence the necessity for more tests. The present writer[1] recently advocated that every child in the elementary school should get at least six hours' testing during his school career, and his only regret is that he did not double this minimum.

The more advanced student may consult Cox's work on the exact measurement of mechanical aptitude. For lack of space no further discussion of the significance of the results for education and industry is here possible. But reference may be made to one urgent problem: Tradition has smoothed the path of the pupil who specialises in pure science. One consequence is that sometimes such a pupil regards himself as an altogether superior being to the specialist in technological science. The truth is often very different. The technological expert often "comes into his own" in spite of the lack of facilities. This is not surprising, for in addition to possessing "g" in high measure, he also possesses a high measure of mechanical aptitude. It is true that he often commands a princely salary but equally true that the State which provides suitable educational facilities for such cases will be amply rewarded.

QUESTIONS

1. Distinguish between tests of mechanical ability and tests of manual dexterity.
2. There is some evidence that a group factor is manifested for boys but not for girls in tests of constructive mechanical ability. Discuss possible explanations (cf. C. Spearman, *The Abilities of Man*, p. 229).

[1] Ll. Wynn Jones, "Vital Issues of Mental Testing", North of England Educ. Conf., 1930.

CHAPTER 18

PERSEVERATION

1. THE REVERSE STROKE TEST

INSTRUCTIONS.—(*a*) "Write 123123 . . . as fast as you can. It will only be for 30 seconds. Do not trouble to write neatly." Before (*a*) is taken Ss should write the figures 123 three times for practice and then draw a line through them so that they will not be counted.

(*b*) "Write 123123 . . . as fast as you can, but this time reverse the direction of your movement in writing each figure. Thus you start each figure at the point where you usually end it and finish it at your usual starting point. Do not trouble to write neatly. If you make a mistake, do not correct it but go straight on without stopping." Time allowed—2 minutes. Before (*b*) is taken Ss should be instructed to start on the left of the paper and go right across to the right, and likewise for the succeeding rows. A blackboard illustration should also precede (*b*) in order to make clear how each figure should be written. Ss should then practice writing 123123 . . . in this way for 15 seconds and then draw a line through them so that they will not be counted. If n_1 and n_2 are the numbers of figures written in (*a*) and (*b*) respectively, then a possible measure of the hindrance in (*b*) would be $4n_1 - n_2$. Another would be $\frac{4n_1}{n_2}$. The symbol ψ, the Greek *psi*, will be used to denote this hindrance or perseveration score, seeing that the letters *p*, *s* and *i* stand for perseveration, secondary function and inertia respectively.

2. THE "IT" TEST

INSTRUCTIONS.—(*a*) "Copy as fast as possible in your own

handwriting the passage given below. Do not trouble to write neatly." Time allowed—1 minute 20 seconds.

(*b*) "Copy the passage again from the beginning, but this time do not dot the *i*'s and do not cross the *t*'s. If you dot one of the *i*'s or cross one of the *t*'s by mistake, do not stop to correct it but go straight on as fast as possible. Do not trouble to write neatly. If you finish the passage before time is up, start again from the beginning." Time allowed—4 minutes. At half-time "Cross" is called out and Ss quickly mark with a cross the point they have reached. As before, possible measures of ψ would be $3n_1 - (n_2 - m)$, or else $\frac{3n_1}{n_2 - m}$, where n_1 and n_2 are the number of words written in (*a*) and in (*b*) respectively, and m is the number of errors (*i*'s dotted or *t*'s crossed).

Paragraph for "IT" Test

Before this account of our interesting climb is finished I must still mention the admirable work of our guide Vincent Smith. I therefore write this tribute to his virile spirit and to the ability with which Smith brought us through the thick mist down the steepest side of the mountain, it was a magnificent exhibition of strength and skill, it was quite impossible to see thirty feet in front through this thick mist yet Smith still went on with the utmost certainty, not once did he retrace his steps or exhibit the least hesitation. At last, though very tired and thirsty, we arrived at the foot of the glacier. Here we met three guides who were setting out to find us, as it was thought we might have met with disaster.

3. SIGNATURES TEST

Until the test is standardised it is suggested that E should ask eight persons whose signatures are rather uncommon to sign their names on a sheet of paper with hectograph ink, so that they may be duplicated. Each S then receives a copy.

INSTRUCTIONS.—(*a*) "Copy this list of names in your own handwriting as quickly as possible. If you finish before time, start again from the beginning." Time allowed—1 minute.

(*b*) "Now copy the signatures as accurately as you can, keeping the size approximately the same." Time allowed—3 minutes.

In marking the test the number of errors is noted. By an error is meant the drawing of a letter in a way which exhibits S's style and which is appreciably different from the original. Measures of ψ would be $3n_1 - (n_2 - m)$, or $\frac{3n_1}{n_2 - m}$, where n_1 and n_2 are the number of letters written in (*a*) and (*b*) respectively and m is the total number of errors in (*b*).

4. MIRROR IMAGE TEST

INSTRUCTIONS.—(*a*) "Write BCDEFGBCDEFG . . . repeatedly for 1 minute as fast as you can."

(*b*) "Write the mirror image form of the same letters as fast as you can." Time allowed—2 minutes. At half-time "Cross" is called and Ss put a cross at the point they have reached. Before (*b*) is taken Ss should practice writing the mirror image form for 15 seconds. $2n_1 - (n_2 - m)$, or $\frac{2n_1}{n_2 - m}$ may then be calculated.

5. "EEE" TEST

INSTRUCTIONS.—(1) "Write eee . . . for 30 seconds as quickly as possible with good quality."

(2) After 5-10 seconds' rest (1) is repeated.

(3) Same as (2).

(4) "Write reverse-stroke *e*'s as quickly as possible with good quality for 30 seconds."

(5) "Write alternately a normal 'e' followed by a reverse stroke 'e' and so on, as quickly as possible for 30 seconds."

A measure of perseveration is

$$\psi = \frac{n_1}{n_2} + \frac{n_1}{n_3},$$

where n_1 is number of letters per 30 seconds when written in the usual way, n_2 is the number of reverses in 30 seconds, and n_3 is the number of normal and reverse letters in (5).

THE CONCEPTS OF PERSEVERATION, SECONDARY FUNCTION, AND INERTIA

More investigations are desirable in order to ascertain the relations between perseveration as operative in tests such as have been described, and perseveration as defined by Neisser, G. E. Müller, Wiersma and others. The psychiatrist Neisser, in 1894, seems to have been the first to use the term "perseveration" to denote the abnormally persistent repetition or continuation of an activity after the activity had been once begun or recently completed.

Müller,[1] shortly afterwards, used it to denote "the tendency which every idea has, after once being in consciousness, to remount freely into consciousness", *e.g.* the humming of a melody or the repetition of a phrase such as Mark Twain's—

> Punch, brother, punch, and punch with care,
> Punch in the presence of the passenger.

Instead of perseveration Otto Gross,[2] in 1902, introduced the terms "primary" and "secondary" functions of the nervous system: "Each nervous element whose primary functional excitement means the occurrence of a presentation in consciousness persists secondarily after the presentation has quitted the span of consciousness. That is to say, it remains for a further long period in a state of after-function". [3]

Wiersma,[4] also Heymans and Brugmans,[5] took over this notion of secondary function and included sensory and motor aspects in their testing. Thus in observing the red and green sectors of a disc rotating at an increasing rate of speed it was calculated that perseverators lost the two colours and commenced to see gray much sooner than non-perseverators. The impression made by the red has already died away with the non-perseverator but not yet with the perseverator, when the green impinges upon the same nervous element. Hence the

[1] G. E. Müller and A. Pilzecker, *Zeitsch. f. Psychol.* Ergänzungsband, I., 1900.
[2] O. Gross, *Die cerebrale Sekundärfunction*, 1902.
[3] C. Spearman, *The Abilities of Man*, p. 44.
[4] E. Wiersma, *Journ. f. Psychol. u. Neur.* 8, 1906.
[5] *Zeitsch. f. Psychol.* 7, 1913.

fusion comes at a lower speed with perseverators. When a similar experiment was demonstrated at the White City Exhibition some years later, it was reported that for Sir William Crookes fusion occurred at a speed of only 900 revolutions per minute, whereas in the case of Mr. Bernard Shaw it was as high as 1800 r.p.m. The former was informed that he was a person who liked to keep working at a problem until it was finished. This apparently was correct as it was only with difficulty that Sir William could be induced to appear punctually at the meal table. Mr. Shaw, on the other hand, was told that he possessed the ability to switch quickly from one topic to another and those who have witnessed his ripostes to hecklers will agree that his powers of repartee excel those of most political gladiators. Be that as it may, it is obviously necessary to employ the method of correlation before any conclusions as to the range of the concept can be formed. Thus the important investigation of Lankes,[1] who included in his tests a variety of different mental activities which were related to the various conceptions of perseveration, gave much guidance as to which activities were intercorrelated.

SPEARMAN'S LAW OF INERTIA.—This law is to the effect that "Cognitive events always both begin and cease more gradually than their (apparent) causes".[2]

Spearman had suggested provisionally that both the well-known factor of "g" and the factor of perseveration are general factors dealing with mental energy; "as 'g' measures its quantity, so the perseveration may express its degree of inertia".[3] It is well to emphasise that the provisional assumption of a "mental energy" is not essential to the discussion. Spearman has now replaced the concept of "energy" by the cognate concept of "power" (energy divided by time).[4] More important is the considerable mass of evidence that perseveration is a broad group factor which varies independently of "g". It is interesting

[1] W. Lankes, "Perseveration", *Brit. Journ. of Psychol.* vol. 7, 1915.
[2] C. Spearman, *The Nature of Intelligence*, p. 133, also *The Abilities of Man*, p. 291.
[3] C. Spearman, *The Abilities of Man*, p. 306.
[4] C. Spearman, "The Theory of 'Two Factors' and that of 'Sampling'", *Brit. Journ. of Educ. Psychol.* vol. 1, 1931, p. 154.

to note that Spearman, even before presenting the Theory of Two Factors in 1904, had concluded that the endowment of an individual for one kind of sensation did not tend to go with that for other kinds to any amount exceeding what is already explained by their respective shares in "g".[1] This negative result is a psychological fact which is, of course, independent of the fact suggested by researches on perseveration, namely, that the *lag* of an individual for one kind of sensation tends to go with that for other kinds.

PERSEVERATION TESTS AND THEIR MARKING

It has been found that tests similar to those described in this chapter intercorrelate. We will assume that perseveration is that factor which partly accounts for this intercorrelation. The nature of this factor may appear in a clearer light if we note that many of the activities of the school child consist of well-established and long-practised habits, and if we ask what occurs when we attempt to introduce a change into such automatisms. Handwriting is evidently an admirable activity for the experiment. Thus in the Reverse Stroke Test, and also in the "IT" Test[2] I have attempted to make S go counter to a strong "set". The "IT" Test is rather different from a test where S would be told, *e.g.* to put a dot over the *e*'s and to cross the *h*'s. This would have involved a new adjustment by addition, whereas in the "IT" Test we interfere with a strong "set" by means of an omission. Such a comparison shows why we cannot conclude that if an individual finds it difficult to break a habit he would also find it difficult to form a new habit. A similar device is employed in the Signatures Test. I am indebted to Professor Spearman for the Signatures Test and have used it in various forms. The present form approximates to Rangachar's adaptation,[3] which seemed to give reliable results. It is, of course, essential that S should not hold the Signatures upside down as

[1] C. Spearman, *The Abilities of Man*, p. 234.
[2] Ll. Wynn Jones, *Perseveration*, Rep. of Brit. Assoc., 1915, p. 698.
[3] C. Rangachar, "An Experimental Study of School Children with regard to Some Racial Mental Differences", M.Ed. Thesis, Univ. of Leeds, 1930. Cf. also "Perseveration among Jewish and English Boys", *Brit. Journ. of Educ. Psychol.* vol. 2, 1932.

a professional forger is supposed to do. The test exhibits great individual differences, some succeed in copying the signatures very well, while others betray their own style of writing in the majority of letters written.

The present form of the Mirror Image Test is due to Bernstein,[1] and the "eee" Test is due to Stephenson.[2]

With regard to the measure of perseveration, or ψ, furnished by each test, Lankes, Hargreaves[3] and the present writer employed the quotient method, that is, the ratio of the efficiency in (*a*) to that in (*b*). Later, Bernstein employed the difference method. Rangachar compared the two methods and found the difference method to yield a higher reliability for each test as well as to furnish higher intercorrelations, but unfortunately the difference measure often correlates highly with speed in the first part (*i.e.* the (*a*) part) of the test, and as will be mentioned later in this chapter, Rangachar employed a measure of perseveration which was independent of speed in the first part of the test.

Another measure of perseveration is that employed by Stephenson in tests like the "eee" Test.

Perseveration and Age.—In 1923 I took three tests similar to those described in this chapter with Standards II.a, III.a, IV.a, V.a, VI. and VII. in an elementary school. It was found that the median perseveration score, or ψ, for each standard was fairly constant between the ages of seven and fourteen.

It is true that Ach[4] later expressed the view that in general the influence of perseveration on the behaviour of the child diminishes with age, but this view is apparently based on the early experiments of Ziehen[5] on the free associations of children and not on any experiments of his own. Moreover, he is careful to specify "*influence* of perseveration" and suggests that it is

[1] E. Bernstein, "Quickness and Intelligence", *Brit. Journ. of Psychol.* Mon. Suppl. No. 7, 1924, p. 13.

[2] W. Stephenson, "Studies in Experimental Psychiatry, 2", *Journ. of Ment. Sci.*, April 1932.

[3] H. L. Hargreaves, "The 'Faculty' of Imagination", *Brit. Journ. of Psychol.* Mon. Suppl. No. 10, 1927.

[4] N. Ach, E. Kühle and E. Passarge, *Perseveration*, 1926, p. 240.

[5] Th. Ziehen, *Die Ideenassoziation des Kindes*, I. Abh., 1898; II. Abh., 1900.

the older child's greater steadfastness of purpose which tends to mask the perseveration.

Pinard[1] has recently come to the conclusion that perseveration actually increases between the ages of eight and twelve. But his measures of perseveration were obtained by the difference method and he admits that the increase may have been influenced by the faster speed at which the older subjects worked.

Perseveration and Race.—Rangachar has recently obtained some interesting results when comparing Jewish school boys in Standard V. of an elementary school in Leeds with corresponding English boys of similar social class.[2] Seven tests of perseveration were applied and for each test the Jewish boys were on the average more "perseverative", as the following table shows:

	$\frac{D}{\sigma \text{ diff.}}$
(1) Inverted S Test	1·51
(2) Alphabets Test . . .	2·80
(3) Reverse Stroke . . .	3·43
(4) Mirror Image . . .	1·12
(5) Signatures	5·47
(6) Triangles	0·76
(7) Capitals	3·72

Rangachar discusses the possible explanation of this difference. It was found that the Jewish boys were on the average a little more "intelligent" as $\frac{D}{\sigma \text{ diff.}} = 1{\cdot}43$ (cf. Chapter 23). But this would not account for it, as it was found that there was no correlation between any of the perseveration tests and "intelligence".

It was also found that the Jewish boys had on the average greater speed in the first part of each test, and as the measure of perseveration, employing the difference method, showed correlation with speed in the first part, it appeared that this would account for the difference. To settle this point, Rangachar re-

[1] W. J. Pinard, "Perseveration and Difficult Children", Ph.D. Thesis, Univ. of London, 1929; also "Tests of Perseveration", *Brit. Journ. of Psychol.* vol. 23, 1932, p. 13.

[2] C. Rangachar, *op. cit.*

marked the tests in such a way that the measure of perseveration was independent of speed in the first part of a test. This was effected by means of Spearman's formula for the correlations of sums and differences. But even now the Jewish boys were on the average more "perseverative". Further experiments along these lines with different racial groups would seem to be desirable. It is interesting to note that Professor Fraser Harris as early as 1908 in his study of functional inertia from a more physiological standpoint states: "Place the Jew, for instance, anywhere under the soft southern sky or the harsh grey north, and he never alters, he is the same physically and mentally as he was a few thousand years ago: neither time nor environment can change him".[1]

RELATION BETWEEN PERSEVERATION TESTS AND CRITERIA OF PERSEVERATION

Lankes framed an interrogatory of thirteen questions which was given to his adult subjects, *e.g.*: "Which would you like better, to pursue one subject of study, one kind of work, thoroughly and fully, or to have several things on hand?" The pool of his tests showed a correlation with this interrogatory of 0·41.

Bernstein estimated the perseveration of school children by careful observations and found a correlation in the neighbourhood of 0·5 between the pool of his tests and the estimates.

Sen Gupta[2] framed an interrogatory of ten questions which was given to two competent teachers with a request that two independent estimates be thereby obtained for every pupil. The estimates of the two teachers showed a correlation of 0·75 (p.e. = 0·04). The pool of his tests correlated only 0·20 (p.e. = 0·085) with the teachers' estimates. The difficulty here is that competent teachers may confuse perseveration with intellectual dullness.

PERSEVERATION AND PSYCHIATRY

There are many clinical observations which would lend colour

[1] D. Fraser Harris, *The Functional Inertia of Living Matter*, 1908, p. 65.
[2] K. B. Sen Gupta, "An Enquiry into Perseverative Tendencies in School Children", M.Ed. Thesis, Univ. of Leeds, 1928.

to the view that melancholics generally are perseverators while maniacs on the other hand would be non-perseverators. Not only that, but the early experiments of Wiersma[1] seemed to offer conclusive proof of it. For example, in observing the red and green sectors of a disc revolving at an increasing rate of speed, melancholics were found to lose the two colours and begin to see only gray at lower speeds than maniacs as shown in the following table:

	Mania (11 Cases)	Normal (9 Cases)	Melancholia (18 Cases)
Revolutions per minute . .	1632	942	732

More recently Professor J. Shaw Bolton, Director of the West Riding Mental Hospital, Wakefield, placed the facilities of his laboratory at my disposal so that tests of perseveration could be applied to the patients. I am also indebted to Dr. M. J. McGrath, the Deputy Director, for his co-operation. As regards the revolving colour discs test,[2] although I adopted improvements due to Lankes, such as descending from a high speed instead of ascending from a low speed, and pasting the colours below the discs so that the subject could refer to them, yet my results did not corroborate those of Wiersma, as will be seen from the following table:

	Mania (8 Cases)	Normal (15 Cases)	Melancholia (5 Cases)
Revolutions per minute (red-green)	443	682	541
Revolutions per minute (blue-yellow)	447	674	483

In other words, the test failed to distinguish between maniacs and melancholics.

As regards motor tests such as have been described in this

[1] E. Wiersma, *op. cit.*

[2] Ll. Wynn Jones, "An Investigation into the Significance of Perseveration", *Journ. of Ment. Sci.*, October 1928. Cf. also "Individual Differences in Mental Inertia", *Journ. of Nat. Inst. of Indust. Psychol.* vol. 4, No. 5, 1929.

chapter the melancholics tended to be perseverators, but even here the results for maniacs did not show a general tendency towards non-perseveration. I also found that fluency tests (cf. Chapter 10) did bear out the general view as to fluency in mania and its lack in melancholia.

Stephenson,[1] using motor tests, as a result of considerable testing found that manic cases involving supervening dementia did not tend to be non-perseverators, but he found other maniacs, of good "g" ability generally non-perseverating. On the other hand, in melancholia and in certain cases of dementia praecox high perseveration was exhibited.

PERSEVERATION AND CHARACTER

Pinard[2] found that about 75 per cent of the most "difficult" and "unreliable" subjects to be extreme perseverators or extreme non-perseverators; and about 75 per cent of the most "self-controlled" and "persevering" subjects showed only a moderate degree of perseveration. But his facts gave absolutely no support to the view that associated extreme perseveration with the several qualities of character which have from time to time been supposed to indicate introversion.

Cattell[3] found some negative correlation between perseveration and the character factor which is usually called Webb's "will" factor or "w". This is in agreement with the previous finding of Lankes. "Low 'w' is associated with both extremes of the perseveration scale, but the correlation of 'w' with 'p' is negative because high perseverators have still lower 'w' than have low perseverators."

QUESTIONS

1. Discuss the possible value of perseveration tests in vocational guidance.
2. Show how you would arrange perseveration tests with special reference to students of typewriting.

[1] W. Stephenson, *loc. cit.* p. 11.
[2] W. J. Pinard, *op. cit. Brit. Journ. of Psychol.* vol. 23, 1932, p. 14.
[3] R. B. Cattell, "Temperament Tests: II. Tests", *Brit. Journ. of Psychol.* vol. 24, 1933, p. 48.

References

1. C. Spearman, *The Abilities of Man*, chap. 17.
2. W. Lankes, "Perseveration", *Brit. Journ. of Psychol.* vol. 7, 1915.
3. E. Bernstein, "Quickness and Intelligence", *Brit. Journ. of Psychol.* Mon. Suppl. No. 7, 1924.
4. H. L. Hargreaves, "The 'Faculty' of Imagination", *ibid.* Mon. Suppl. No. 10, 1927.
5. Ll. Wynn Jones, "Individual Differences in Mental Inertia", *Journ. of the Nat. Inst. of Indust. Psychol.* vol. 4, No. 5, 1929.
6. H. M. Cushing, "A Perseverative Tendency in Pre-School Children", *Arch. of Psychol.* No. 108, 1929.
7. J. Seemann, "Untersuchungen über die Psychologie des Rechnens und der Rechenfehler", *Arch. f. d. ges. Psychol.* 69, 1929. (Perseverative errors were found to be the most frequent of all forms of errors in the fundamental operations of arithmetic.)
8. C. Rangachar, "Perseveration among Jewish and English Boys", *Brit. Journ. of Educ. Psychol.* vol. 2, Part 2, 1932 (adapted from M.Ed. Thesis, Univ. of Leeds, 1930.)
9. J. W. Pinard, "Tests of Perseveration", *Brit. Journ. of Psychol.* vol. 23, Parts 1 and 2, 1932.
10. W. Stephenson, "Studies in Experimental Psychiatry: I.—A case of General Inertia. II.—Some contact of *p*-Factor with Psychiatry. III.—*p*-score and Inhibition for High-*p* Praecox Cases", *Journ. of Ment. Sci.*, October 1931, April 1932 and October 1932.
11. R. B. Cattell, "Temperament Tests", *Brit. Journ. of Psychol.* vol. 24, Part. 1, 1933.
12. W. Stephenson, " An Introduction to so-called Motor Perseveration Tests", *Brit. Journ. of Educ. Psychol.* vol. 4, Part 2, 1934.

CHAPTER 19

PRACTICE, FATIGUE AND OSCILLATION

MATERIALS.—Kraepelin's Adding Sheets. Stop-watches.

DIRECTIONS.—S is directed to work down the columns of single-place numbers on a Kraepelin's Adding Sheet, adding each number as it is reached to the number immediately preceding it, and writing down the sum of the two numbers, omitting the left hand digit when the sum is greater than 9, as in the following example:

7	
2	9
8	0
5	3
3	8

Each column of figures should be started afresh, the top figure in any column not being added to the bottom figure in the preceding column. The work of adding should proceed continuously as quickly as possible for ten minutes. Students work in pairs, each in turn acting as E and as S. E should place himself so that he can see S's work over S's left shoulder. E should have a Kraepelin Adding Sheet resting on thick cardboard in his left hand. A stop-watch should also be firmly fixed to the cardboard at the most suitable point for E's vision. E should draw a horizontal line every five seconds at the points on his own sheet corresponding to S's progress. These lines should be drawn noiselessly so that S cannot ascertain when it is done.

It is recommended that this experiment should be repeated on several days.

DISCUSSION

There is a common impression that if an activity be repeated there would be an initial facilitation due to practice followed

by a deterioration due to fatigue. The assumption being that practice is in operation at the start and that fatigue comes into operation at a later stage. Closer study leads to the view that both practice and fatigue have to be considered from the very beginning.

Mental work by occurring tends to reoccur more easily. That is the practice effect. Mental work by occurring produces a tendency opposed to its occurring afterwards. That is the fatigue effect. There are thus two contrary factors at work, and as regards these two factors the resultant will depend on the relative intensities. A study of the work curve may give some indication as to which factor gains the upper hand at various points of its course.

But there are many other factors besides practice and fatigue which must be considered, such as the "warming up" of the subject or, again, the "spurts" which he may make at any point of the work. The entry of conative and emotional factors, however, affect the shape of an individual's work curve in real life to so great an extent as to necessitate other methods of study in addition to laboratory studies. Industrial unrest, monotony, a hard taskmaster, the incentive of a "glittering prize" and many other factors exert a variable and uncertain influence on the work curve, and the industrial psychologist attempts to improve the shape of the work curve by studying the factors which affect it adversely. Thus the introduction of a suitable rest pause may not only increase the output but may also engender a more favourable attitude of the employee towards his work. Similarly, attention to the lighting, temperature, humidity of the atmosphere, ventilation, general arrangement of the machinery, posture of the workers, and so on, have an obvious bearing on the work curve.[1] In this way by controlling the conditions, industrial psychologists have been enabled to study the work curve, complex though it be; and although such a study need not isolate "fatigue" or attempt its definition, it is yet clear that it may materially diminish it.

Many laboratory investigations, however, have been devoted to the study of fatigue, and as the difficulties are great, mention

[1] C. S. Myers, *Industrial Psychology in Great Britain*, 1925, chaps. 2 and 3.

will only be made here of some general results. In the first place, it is necessary to distinguish between subjective and objective fatigue:

SUBJECTIVE FATIGUE.—Subjective estimates of the state of fatigue such as "very tired", "tired", "fairly tired", "slightly tired" or "not at all tired" are not difficult to obtain. Phillips,[1] in this way, was able to conclude that the person who experienced most fatigue in one task tended to experience most fatigue in others. In statistical language this was expressed by saying that the estimates of subjects as to their state of fatigue for four fatigue-producing tasks were positively correlated and that the tetrad-difference criterion (see Chapter 23) was satisfied. Subjective fatigue, then, is general in the sense that the individual most susceptible to fatigue for one task will tend to be most susceptible for others also. Such a result would appear to agree with common observation.

Subjective fatigue, however, was not found to be general in respect of transfer, and this result again is in agreement with the fact that individuals after becoming tired at one task often lose that tired feeling on changing to another. But that is not to say that fatigue of an objective nature was not present.

OBJECTIVE FATIGUE.—Thorndike has defined objective fatigue as "that diminution in efficiency which rest can cure", and there is considerable evidence that although a subject's actual efficiency in a given task may decrease but little after the onset of objective fatigue, yet the longer the task is continued, the longer will be the rest necessary in order to regain full initial efficiency. This rest then becomes a measure of objective fatigue, and it renders patent what would otherwise remain latent.

Many of the methods proposed to measure objective fatigue are open to criticism. They may be divided into two classes:

(1) Sensory and Muscular Tests, *e.g.* the Aesthesiometer Test, the Ergograph Test, or the Reaction Time Test.

(2) Mental Tests, such as Adding Figures or Cancellation of Letters.

[1] G. E. Phillips, *Mental Fatigue*, 1920, p. 66. Cf. also C. Spearman, *The Abilities of Man*, p. 314.

Such methods must be criticised if they assume without proof that the fatigue which may have been produced from a particular activity is transferred to the activity which is demanded by the test. On this matter the work of Phillips is the most illuminating: "If a child is found to be normal in a certain test at the end of the school day, it does not follow that he is not fatigued. The only inference to be drawn from such a result is that he is not fatigued *in that particular test*; but the result gives no evidence as to his state of fatigue in any other test, nor does it give any evidence of the state of fatigue in any test of any other child who has done exactly the same work as the former child. If two subjects, after the same duration of mental work, are tested, say, in multiplication and cancellation, the first may be very fatigued in multiplication, and only slightly affected in cancellation, while the results for the second subject may be the exact reverse. He may be very fatigued for cancellation, and only slightly fatigued for multiplication."[1] Such considerations show that the earlier experiments on fatigue due to school work cannot be regarded as valid. It is more profitable to study the variations in output during school hours. A similar method has proved serviceable in the study of industrial fatigue. Output and its fluctuations are measured, then changes in the lighting, in rests, in wages, etc., are introduced with a view to increasing the efficiency, and it is not strictly necessary to assume that fatigue has been diminished.

Dawson[2] studied the variations in the mental efficiency of children aged eight to fourteen during school hours. The task was to multiply figures for ten minutes at different work-periods, namely, 9.30, 10.30, 11.30, 1.30, 2.30 and 3.30, and the number of correct operations and errors were calculated. The output was approximately constant from 9.30 to 2.30. It is true that an increase of output of 8 per cent was found to occur at the 10.30 period, but this, on analysis, was not found to be due to its being the best working hour of the day as popularly supposed but to the fact that this is the usual arith-

[1] G. E. Phillips, *op. cit.* p. 74.

[2] S. Dawson, "Variations in the Mental Efficiency of Children during School Hours", *Brit. Journ. of Psychol.*, 1924.

metic hour. A habit-attitude had been developed which favoured the performance of arithmetical operations at this hour. As the 3.30 period—unlike any other—came last for each group the increase of output which occurs is due to practice, and further experiments are necessary to ascertain if there would be a falling off in efficiency at this hour if the effects of practice had been eliminated. It was suggested however, that for children below the age of nine, there is a distinct falling off in efficiency by 3.30 in the afternoon.

FATIGUE AND LOSS OF SLEEP.—Miss M. Smith[1] carried out an important study of the effects of curtailing the normal amount of sleep for three successive nights. The normal amount was 8 hours, but on the first night's vigil only 1½ hours' sleep were taken, on the second 3½, and on the third 5½ hours'. On testing the subject by means of McDougall's Dotting Machine and other tests, it was found that the immediate result of loss of sleep is an increase of efficiency which then falls below normal when the usual number of hours' rest is taken. If the usual amount of sleep is taken until the efficiency again reaches the normal, it was found that this point is not reached until about the sixteenth day from the beginning of the experiment. Possibly the initial phase of the fatigue is the production of some toxins which have a stimulating action like drugs which is followed later by a period of decrease in function and depression. Interesting as these researches are, it is necessary to ascertain what differences there may be when several subjects are tested before general conclusions can be drawn.

THE INFLUENCE OF DRUGS.—Careful experiments by Dodge and Benedict,[2] and also the later experiments of Hollingworth,[3] show that alcohol has an adverse effect on mental and motor efficiency. It is necessary in such experiments to follow Rivers's[4] lead and disguise the alcohol so as to provide a con-

[1] M. Smith, "A Contribution to the Study of Fatigue", *Brit. Journ. of Psychol.*, 1916.

[2] R. Dodge and F. G. Benedict, *Psychological Effects of Alcohol*, Carnegie Institute of Washington, 1915.

[3] H. L. Hollingworth, "Psychological Influence of Alcohol", *Journ. of Abn. and Social Psychol.*, 1923 and 1924.

[4] W. H. R. Rivers, *Influence of Alcohol and other Drugs on Fatigue*, London, 1908.

trol and make it impossible for the subject to know whether it is present or not. The disastrous effects are proportional to the size of the dose. The effects of other drugs such as strychnine, morphine, opium and cocaine are in general an initial stimulation followed by a period of depression. It is therefore highly probable that their habitual use would be likely seriously to undermine the general health.

Practice, Fatigue and Oscillation.—Kraepelin's Adding Sheets were used by Flugel[1] in his study of the factors of practice, fatigue and oscillation. 46 girls from Standards VI. and VII. worked at these sheets for 20 minutes on each of 46 successive school days. Strong incentives were employed in order to ensure that the work was done with maximal effort. All the subjects were keen and interested and kept graphs so that they could vividly realise the nature and extent of their progress. Each girl received a monetary reward, and every time she exceeded her previous record an extra bonus was given in addition. Moreover, Flugel originated the novel method of increasing the amount of this bonus as the experiment proceeded, for it was less easy to improve in the later stages. In consequence the curve of practice showed continuous improvement throughout the experiment. The composite curve of practice was obtained by plotting the total work done by all the girls for each of the 46 days. It was found that the total score on the last day was about $3\frac{1}{4}$ times as great as that on the first day. Further, by plotting the work done by each girl during the first, second, . . . minute of work the curve of fatigue was obtained. By adding the work done by all during each minute a composite work curve was obtained. A marked drop in efficiency occurred at the beginning of the first minute and thereafter fatigue manifested itself as a slow and more or less steady decline throughout each day's experiment. This rapid initial decline is a general characteristic for all subjects, but considerable differences between individuals as regards fatigability were found. It is to be noted that in Flugel's tests a bell pealed every five seconds and the subjects drew a short line

[1] J. C. Flugel, "Practice, Fatigue and Oscillation", *Brit. Journ. of Psychol.* Mon. Suppl. No. 13, 1928.

under the figure last written. In addition, at the end of every minute "Now" was called and the subjects drew instead an L-shaped line. Further, each day's work commenced with a "flying-start", inasmuch as 13 five-second periods were allowed to elapse before the first one-minute signal was given. The work done in the first of these 13 periods was disregarded.

Flugel points out that the giving of the one-minute signals, and possibly also the giving of the five-second signals, introduced an "artificial rhythm" which prevented him from studying any "natural" rhythms in the work curve. It would be difficult to avoid giving such signals in a group experiment. Flugel's aim in dividing the work into five-second periods was to study the short-time fluctuations in efficiency in the work of his subjects. The difference between the scores of each successive five-second period was calculated for each subject. The sum of the differences obtained for any subject provided a measure of his total tendency to oscillation. A measure of the mean oscillation was obtained by dividing this total oscillation by the total number of calculated differences between the scores of the successive five-second periods. Finally a measure of percentage oscillation was obtained by expressing the subject's total oscillation as a percentage of her total score. Increase of the ability of adding through practice was accompanied by an increase (though not a proportionate increase) in the extent of oscillation. But decrease of ability through fatigue was not accompanied by any corresponding decrease of oscillation. As fatigue increased the subjects tended to oscillate more. This suggests that oscillation is a phenomenon which is derived from fatigue.

There is some evidence[1] that oscillation is a new factor of wide extent and that an individual who tends to oscillate in one work test, such as that performed on Kraepelin's Adding Sheets, will also tend to do so in other work tests such as subtraction, cancellation, opposites, or analogies. Recent work by Sivaprakasam, however, appears to show that the oscillation

[1] Cf. C. Spearman, *The Abilities of Man*, p. 324, and K. Sivaprakasam, "Oscillation of Attention", Ph.D. Thesis, Univ. of London, 1933 (on file at the Univ. of London).

in the reversible perspective test is specific to the test and not that which is common to the work tests. He, moreover, concludes that the oscillation factor common to the work tests depends largely on reproductive rather than on eductive ability.

Some of the formidable difficulties inherent in an experimental study of fatigue may at this stage be summarised:

(1) *The Control of Physical Conditions.*—In testing intelligence investigators do not impose any prandial restrictions on the testees, but the writer well remembers a foreign investigator having to meet the opposition of some of the parents of his subjects when he asked them, quite rightly from his point of view, to take his tests of fatigue before breakfast. In addition to food and drink, there are many other factors to be kept under control, such as ventilation, humidity, illumination and the shape and size of chair and bench.

(2) *The Elimination of Emotional Factors.*—Disturbances of this nature may not play an appreciable part in the laboratory, but when industrial psychologists attempt the investigation of fatigue from the study of output they do not neglect to ascertain if suspicion and ill-will may be ruled out, for they may seriously affect the results.

(3) *The Control of Conative Factors.*—There is, perhaps, a natural but undue tendency to assume that most subjects, when psychologically tested, are performing the tasks with concentrated effort. Or, at least, to assume that the slackers, even when groups are tested, are too few appreciably to affect the results. But the experiments of Flugel show how strong incentives may affect the work curve. His subjects seem to have shown throughout the investigation a high degree of conative activity. But in daily life there is not always a glittering prize or infuriated bull to bring about maximal efforts. At least part of our work is done at the time we choose and at the rate we find suitable. Enlightened employers do not expect maximal effort, except for brief periods. Slave-drivers may expect it, but they will seldom get it.

(4) *Diurnal Variations of Efficiency.*—There are probably diurnal variations in efficiency which are independent of fatigue. Unless their magnitude and location are known, they may

conceivably render invalid certain conclusions as to the presence and degree of fatigue.

(5) *Objective and Subjective Fatigue.*—A distinction has been made between a loss of ability for an operation and a loss of inclination for it. "The objective kind cannot, but the subjective kind can, be banished by emotion or overcome by will."[1]

(6) *The Fatigue Indicator.*—Unless it is true that objective fatigue is tranferred from one task to another independently of the nature of either task as well as independently of the subjects who do the task, it is obvious that no test can be considered as an adequate measure or indicator of the fatigue incurred with every kind of work. The difficulty is increased owing to the peculiar course taken by fatigue in its development. Investigation has shown that, "on eliminating irrelevant influences, a person's efficiency at any continuous work diminishes very rapidly for a period of about two minutes, but then undergoes very little diminution for hours, until finally there is an abrupt drop down to entire impotence. But during all the time that so little diminution is manifested, a longer and longer rest becomes necessary in order to regain the full efficiency of the beginning: fatigue was really occurring all the time, but in such wise as to remain latent."[2] Such considerations will throw light on the results of such experiments as were carried out with subjects who pedalled a laboratory bicycle. Although there was no apparent and objective indication of fatigue, yet after some hours the subjects collapsed.

Fox[3] regards the interference with steady or rhythmical activity as the essential phenomenon of fatigue. On calculating the coefficient of variation (cf. Chapter 23) of the marks for each subject, for every consecutive five-minute period of a task which lasted thirty minutes, he concluded that during the last ten minutes the work was performed more irregularly. It is interesting to relate this to Flugel's tendency to oscillation which appears to be derived from fatigue. But from what has already been written in this chapter it is obvious that we cannot

[1] C. Spearman, *The Abilities of Man*, p. 309.
[2] *Ibid.* p. 309.
[3] C. Fox, *The Mind and its Body*, 1931, p. 172.

regard this lack of steadiness as the *only* indication of fatigue. Moreover, Fox *assumes* that the irregularity during the first five-minute period is due to his subjects not having adjusted themselves to the task, and it may be noted that this initial irregularity is far greater than any subsequent irregularity.

(7) *Individual Differences.*—Owing to the difficulty of eliminating irrelevant influences, little is known about the various ways in which individuals may differ from each other in fatigability.

(8) *Theories of Fatigue.*—It would be outside the scope of this book to refer to biochemical, physiological or neurological considerations which bear on theories of fatigue. It may suffice to mention three causes of fatigue without attempting to assess the relative weight of each in a particular case.

Inhibition Theory.—Briefly stated, it is held that when one continues to contract a muscle voluntarily, as would be the case in a monotonous task not necessarily involving heavy toil, inhibitory nervous impulses ascend from the muscle to the central nervous system. In consequence it becomes more and more difficult for impulses to descend to make it contract.[1] Eventually the work stops, and there is an intense feeling of boredom.

Toxin Theory.—The activity of a muscle produces waste products such as lactic acid and carbon dioxide and until these are dissipated they act as poisons and hinder further activity.

Fuel Consumption Theory.—All activity uses up definite materials such as glycogen, and it has been assumed that fatigue is the result of an excessive use of these materials which can only be made good by the organism during its rest periods.

Psychologically the upshot would seem to be that if we continue at a muscular or mental task, it eventually becomes distasteful and we experience fatigue. Unfortunately, however, considerable objective fatigue may be present without its being experienced by the individual. He is then unable to pay heed to this natural safeguard against injury.

On the other hand, there are occasions in actual life when work has to be done during a state of fatigue. Whether due to

[1] C. S. Myers, *Industrial Psychology in Great Britain*, p. 42.

the inhibitory action of the central nervous system, or to the production of toxic products, or to the using up of fuel is for the moment immaterial. Assuming that the individual is organically sound, he may by judicious training come to acquire the attributes of tempered steel. Thus anyone electing to compete in a half-mile race without previous preparation would experience extreme physical discomfort and feelings of fatigue. After training, the discomfort and fatigue although present in a lessened form, may be controlled and the runner's "second wind" may enable him to win. Similar considerations apply to mental tasks, such as those of a successful barrister. But in each case, if a breakdown is to be avoided, it is necessary to use considerable judgment in applying the principle.

QUESTIONS

1. What conclusions may be drawn from the study of an industrial work curve? (Cf. C. S. Myers, *Industrial Psychology in Great Britain*, p. 68.)
2. Contrast the attitude of a subject in the laboratory who undertakes continuous work with that of an industrial worker at his daily task. (Cf. C. S. Myers, *ibid.* p. 48.)
3. What is meant by an individual's tendency to oscillation? Show how it can be measured. (Cf J. C. Flugel, *Practice, Fatigue and Oscillation*).
4. In what respects is subjective fatigue general? In what respects not general? (Cf. C. Spearman, *The Abilities of Man*, chap. 18).
5. What is meant by the statement: "Subjective fatigue is wholly specific, whilst objective fatigue is partly specific and partly general"? (Cf. C. Spearman, *ibid.* chap. 18).

REFERENCES

1. G. E. Phillips, *Mental Fatigue*, publ. by W. A. Gullick, Sydney, 1920.
2. C. S. Myers, *Mind and Work*, 1920, chap. 2.
3. C. S. Myers, *Industrial Psychology in Great Britain*, chap. 2, 1925, 2nd and rev. ed., 1933.
4. C. Spearman, *The Abilities of Man*, chaps. 18 and 19.
5. F. Aveling, *Directing Mental Energy*, 1927, chap. 4.
6. J. C. Flugel, "Practice, Fatigue and Oscillation", *Brit. Journ. of Psychol.* Mon. Suppl. No. 13, 1928.

CHAPTER 20

MARKING

MATERIALS.—Typed copies of the seven compositions on "A Wet Day", referred to as Nos. 19, 24, 22, 15, 10, 2 and 26, written in 30 minutes without previous preparation by Standard VII. boys.

DIRECTIONS.—Copies of Nos. 19, 24, 22, 15, 10, 2 and 26 should be handed to each member of the class After E has explained what they are, each member is requested to read them carefully and to award each a mark, 10 marks being the maximum. As the essays are typed, judgment is not affected by the quality of handwriting. Further, the more obvious grammatical errors have been corrected. After the marking has been completed, E exhibits a table on the blackboard showing the marks given to each essay by every member in the class. Then the marks given to each essay by every member are added up, and hence the average mark for each essay may be calculated. Each member then adds up the differences between the mark he has given to each essay and the corresponding average mark, and divides this by seven to obtain his average deviation. Which member of the class agrees best with the composite marks? If the composite marking be accepted as more reliable than that of any individual in the class, who is inclined to mark too leniently, or too severely?

Each member of the class should then write an account of the method which he employed in marking. E should then collect these accounts and initiate a discussion as to the methods which have been employed. It will then be seen that different individuals employ widely different criteria in marking compositions and stress some criteria more (or less) than other individuals do.

No. 19

It was Saturday morning and the night before I had gone to bed

with the idea of going to the football match on the Saturday afternoon, but to my disgust it was raining in torrents. To make matters worse, it started thundering and lightning. It was impossible for me to go out and I had to sit near the window. The football match would be abandoned and people like myself would be disappointed and very likely they would have to stay at home, or perhaps they would go to the afternoon pictures. Time passed on, and then dinner time, but I had not much thought for that. I was just finishing my dinner when my pal Mason came to see if I would go to the pictures with him, and after considering for a moment I said "Yes, you can come in and wait for me". Soon we were walking down the street, not too fast for fear we should slip. We had not been out many minutes before we were soaked to the skin. The roofs were shining with the rain and it was still running off. The clouds were black and they were travelling at a terrific pace. When I came from the pictures it had stopped raining and a big rainbow was in the sky.

No. 24

I woke up one morning to see a dark, dismal world. Outside I could hear the rain falling in torrents, and the wind howling and whistling. Here was a nice thing, no strolling through the woods, no playing on nice green turf for me. I got out of bed and switched on the light, slowly I washed and dressed and went downstairs. There I met my sister, "Lovely day, isn't it?" she said. I surveyed her gloomily. "You look cheerful, sorry you got up." "Oh, shut up." She shut up accordingly. I began to make the fire, laid the paper and wood, went for the coal. When I came back the paper was soaked! My sister had been writing, she looked up and burst out laughing at my plight. "I'll get you some more paper", she said, and went for some. I saw what she had been writing: "For a person who has the grumps". There was a list of things a wet day does: "Makes grass fresher", well, so it does. "Makes dust disappear". Good, so it did. "Makes a walk better." So it does. Hurray! After all, a walk in the rain is fine. Good boots on, a mackintosh on, and some [1] in my mouth. Fine!

No. 22

A wet day has disadvantages and also advantages. In the first paragraph I am going to say a little about its disadvantages.

(*a*) There are many poor people who have not a pair of boots to wear. There are others whose only roof is the sky if you are out in

[1] Play on words, reference to toffee.

the country and the rain comes down in a deluge. These are only one or two ways in which a wet day has its disadvantages.

(*b*) I will now say a few words about its advantages. When it rains it washes the flags and roads. It also draws up the flowers and trees from their beds of earth. If a person feels glum, he just wants to put his hat and coat on and go for a walk. Rain is a good medicine, it cures people who are miserable. If it were not for wet days we could not live. It helps wheat, potatoes and all the food we eat. The rain provides us with a beverage, but before we drink it, it must be purified. Rain plays a great part in making the beauty of the world. It also affects sports. A wet day is one of the greatest gifts of God. Most people grumble when the sky is dull and grey.

No. 15

On a very wet day the rain can be heard banging and rattling on the glass windows. The gutters are filled with water flowing quickly to the grates. If you are caught in the rain when out in the country, you generally get a cold if you have not good boots on.

In the country when it is raining, the ground is very muddy. The wind blows strongly and the streams flow fast. The rain has advantages as it keeps the dust down, washes the flags and the roads, and waters the garden. The reservoirs are also filled. There are a few indoor games which you can play on a wet day, such as Ludo or snakes and ladders, cards, solitaire and draughts. There are some people who like to go out when it is raining and have a good walk. In some of the country houses where there are no taps, the owner puts a big barrel under the fall pipe to catch the water which is used for washing. The rain comes sometimes when it is not wanted. If you have arranged to go for a picnic and it starts raining you get very disappointed.

No. 10

On a wet day the sky is very dull which sets a fog on to the earth below. The clouds are very low and dull, everything seems to be one huge dull mass of atmosphere. Then when the clouds get to the mountains, one by one they burst, and tons of rain pour forth to the solid earth below. Some people like going out in the rain and dampness of the country and lanes. Other people who do not like this stay indoors and read, or, if they do not like this, they sit and watch the rain pattering on the flags and windows. Most children play games which they have such as Ludo, snakes and ladders, or probably they play billiards if they are so lucky that

they have a billiard table. The scenery looks dismal and damp, water is running down the gutter as fast as it can. Sometimes when it rains, it starts lightning and thunder. Street lamps are reflected on the pavements and drain pipes are wet with rain pouring down them. In the country everything looks desolate and forlorn. There seems to be no living thing at all about for miles around. When it stops raining and the sun starts to shine forth its beautiful rays of warmth and light, everything changes and instead of being desolate it becomes beautiful and light. It is like coming into a new world. Children bring their bats or footballs out on their moors or recreation grounds for outdoor sports.

No. 2

The hot, dry, gloomy day was coming to a close, one could not help feeling a little depressed. As soon as it became dusk I noticed a suspicious moisture in the air. I felt greatly relieved when it started to rain. It felt cool and refreshing as it came lightly on my face. It was now dark and the lights began to appear. All the gloominess had been chased away by the little glittering raindrops. The lights were reflected on the pavement, the motor headlights seemed to dance up and down like elves. Everything seemed bright and alive. One did not feel miserable, but (for once) went about with a cheerful face. The roofs were shining, the pavement was glittering. It was not heavy rain but came down lightly. Even urchins, most of whom were barefooted, did not heed the fine and now penetrating rain but still played with their peculiar games as usual. Now people seemed to "fall off" and go to their sheltering homes, for one can have too much of a good thing. People were getting thoroughly wet, they were beginning to grumble and say that it was just like the English climate, it was either one thing or the other and there was no medium. The streets became desolate and the few people who were about had that bored expression, like they had in the earlier part of the day.

No. 26

On wet days some people are miserable. Other people are not so miserable. The rain makes the roof glitter. The rain makes little pools in the hollows. A wet day makes the roads and lanes very muddy and dreary. When the motors and other things go down in the hollows their sides get muddy. It sometimes makes the people's clothes very muddy. The rain makes the lamps very clean. The rain bounces off the window sills. It runs off the ends of the roofs, and it runs down the pipes. It also runs in the gutter and it runs down

the grates. On a wet day the rain gets in the cracks in the flags. It makes the flags and road very clean and brown. It also makes the slates on the roofs very clean. On a wet day the sky is covered with clouds. The rain makes a patting sound on the windows. It leaves mud in the gutters.

DISCUSSION

To award marks is a very ancient human characteristic. Many centuries ago candidates for the Chinese civil service were confined in the examination rooms for many days while they wrote down all they had memorised from books; and the marks which they received were probably fairly reliable for picking out the best memorisers whatever their efficiency might be for choosing the best civil servants. At the present day marks are of enormous importance in attaining many positions and honours, excepting those which depend on social standing, family relationship, fraternity influence, political or religious views and so forth.

Until recently it was fairly generally believed that marks were very reliable measures; but it obviously depends in part on the activity marked. It is easier to mark a Bisley marksmanship test than it is to mark a geometry paper. A facsimile reproduction of a pupil's examination paper in plane geometry was sent to 115 schools with a request that it be marked by the teacher of geometry. Of the 115 marks, two were above 90 (out of 100), one was below 30. Twenty were 80 or above, while twenty others were below 60. The passing mark was 75. Forty-seven teachers assigned a mark passing or above, while 68 teachers thought the paper not worthy of a passing mark.[1] Thus the mark which a pupil receives depends upon the teacher who marks as well as upon what the pupil places upon the paper.

A more difficult feat, however, is to mark English compositions. As a rule the teacher has a definite aim in setting a composition and marks it accordingly. In the foregoing experiment, however, it was thought expedient not to give any explicit directions as to the criteria to be employed in marking,

[1] D. Starch and E. C. Elliott, "The Reliability of Grading Work in Mathematics", *School Review*, vol. 21, 1913. Also quoted in D. Starch, *Educational Psychology*, 1920, p. 434.

but rather to find out first of all how widely they differ with the individual.

I ascertained what criteria were employed by members of my classes, but this is now purposely omitted as it might introduce some bias into the marking. I once requested a class to mark a series of twenty-nine essays (actual copies) but to pay no regard to the quality of handwriting or to spelling errors. These essays were then typed after the spelling errors had been corrected. When the same students marked the typed copies a fortnight later, it was found that notwithstanding the instructions given, some had nevertheless been influenced in the previous marking by the relative poorness of the handwriting or spelling with a consequent raising of the marks for such cases in the second marking.

It is not, of course, suggested that bad penmanship or bad spelling should be disregarded in marking essays. It is rather a question of ensuring that too much stress is not placed on these factors to the detriment of other factors. As marking has such important consequences for the community it is necessary that markers should be trained[1] to mark justly and efficiently. "To err is human," but with care many marking errors can be diminished if not eliminated. Thus for certain examinations involving hundreds of examinees, a perfectly simple statistical check is placed on the marking of the university teachers who act as examiners to ensure that the standard of marking does not appreciably change from hour to hour. Just as the sense of humour, politeness or impeccability of a pupil may affect our estimate of his intelligence, so may our own changeable moods or surroundings affect our estimates of our fellow-men and also of such products as compositions written by persons entirely unknown to us. When, however, the writers of the compositions are known to us, other complications arise, as is shown by the following instance well known to the writer.

A dozen members of a Fourth Form decided to play a practical joke at the expense of their teacher. Each wrote an essay which was then passed to another, who copied it into his own

[1] Cf. D. Speer, *An Experimental Evaluation of Seven Composition Scales*, The John Hopkins Univ. Studies in Education, No. 14, 1929.

class book and handed it in to be marked. The result showed conclusively that this particular teacher marked individuals and not essays, for the pupil who was accustomed to get a good mark still got it even when his essay was composed by one accustomed to receive a low mark. Moreover the one who was accustomed to a low mark did not get a high mark even when his essay was composed by a brilliant pupil! It is not suggested that such a result would ensue universally, but a comic instance of a similar kind has been well described by Ballard.[1]

If a number of examiners had to mark a hundred essays on a given topic written by a random group of Standard VII. boys, it would be instructive to ask each to place the essays in five boxes labelled A (very good), B (good), C (medium), D (poor) and E (very poor). If the distribution is approximately normal the biggest number would be found in C, and the numbers in A and E would be small. If now those in each box are further analysed, it will be possible to pick out the very few in A which are especially good and they may even be given the maximum of 10 marks, all others in A receiving 9 marks. Similarly a number in B may be awarded 8 marks while the majority receive 7. Those in C may be divided into two roughly equal classes to which the marks 6 and 5 are awarded. Those in D get 4 or 3, those receiving 4 being in the majority. Likewise in E 2 or 1, those receiving only 1 mark being probably very few in number. In this way the distribution of each marker tends to be normal, but if each examiner marked by general impression such a method is obviously very subjective in character and wide differences between the markers would be expected, even when the examiners are acquainted with the various qualities of essays to be expected from Standard VII. boys.

The following table shows the distribution of marks given by three students (purposely picked out from a larger number) when marking the twenty-nine essays already mentioned:

Marks . .	1	2	3	4	5	6	7	8	9	10
Student X .	0	0	1	3	12	8	3	1	1	0
Student Y .	0	0	0	0	3	12	5	6	2	1
Student Z .	0	0	2	7	7	3	6	3	1	0

[1] P. B. Ballard, *The New Examiner*, 1923, p. 54.

Student X has the most symmetrical distribution. Student Y awards half marks or more to all the boys. Nine boys out of twenty-nine receive less than half marks from Student Z.

The credit for the first serious study of the statistics of examinations seems to belong to Professor Edgeworth who analysed the various errors involved in marking, and ascertained the probable magnitude of each in various types of examinations. A brief summary of his work is found in Hartog's *Examinations and their Relation to Culture and Efficiency*.

More recently Starch[1] and others have studied the statistics of school marks. Starch mentions four possible factors to account for differences in marking: (1) Differences in severity in different schools. (2) Differences in severity of different teachers. (3) Differences in credit for a given question or in penalty for a given error assigned by different teachers. (4) Minuteness of discrimination between successive steps of merit—the *minimum sensibile*.

From a study of the mean variation of the marks he concluded that factor (1) might almost be neglected, factor (2) accounted for approximately a fifth, factor (3) accounted for nearly a half, and factor (4) for nearly a third of the total mean variation. He thus concluded that factors (3) and (4) are the most important in producing the large differences of values assigned by teachers to a given piece of school work.

It is astonishing how persons of academic standing differ from each other when they mark the essay type of answer. Some will only give a low mark if it contains one or two literary blemishes which they dislike, others award marks for knowledge displayed and pay but little heed to the style in which it is presented. The situation is serious when it is considered how frequently an examination question begins with one of the following words: analyse, compare, consider, contrast, criticise, describe, discuss, distinguish, estimate, examine, explain, give, illustrate, indicate, outline, show, trace. Knowledge and style are both important. Happy is the examiner who can weight each as the situation demands!

[1] D. Starch, *Educational Psychology*, 1920, p. 435.

Burt[1] points out that the common method of marking compositions where positive excellencies are ignored and definite faults are counted and the total subtracted from an arbitrary maximum is almost worthless. He gives valuable hints to the teacher how to form a schedule of items including: (1) the more mechanical aspects—writing, spelling, punctuation, grammar, syntax; (2) the more literary aspects—range, correctness and appropriateness of information, of vocabulary and of rhetorical devices; and above all (3) the logical aspects—the general organisation of ideas, as revealed by the unity, the complexity, the relevance and the sequence of sentences and paragraphs, and indeed, by the intellectual structure of the essay as a whole. Such advice would materially help pupils to acquire the art of writing, but the necessity for such advice shows in a clear light the difficulty of the marker's task in examinations of the essay type.

Ballard[2] has pointed out the grave defects of the essay as a measuring device and concludes that the difficulties of marking an essay are almost insuperable, and refers to the impossibility of making the essay amenable to rigid objective measurement.

Godfrey Thomson and Stella M. Bailes[3] also draw attention to the fluctuating and subjective nature of a judgment on an essay.

It is obviously necessary to ascertain what degree of agreement there would be among different judges marking a set of essays. The evidence available seems to point out the necessity of employing more than one examiner.

There is also the need of comparing scientifically various forms of the analytic method where several virtues are weighted and the general impression method. Each method has adherents.

One writer[4] recommends the impression method owing to the possibility of error in the weighting of the virtues and in the evaluation of each essay for each virtue. Twelve grades are recommended, namely, A. B, C, D, with plus and minus values

[1] C. Burt, *Mental and Scholastic Tests*, p. 330.

[2] P. B. Ballard, *op. cit.*

[3] "The Reliability of Essay Marks", *Forum of Education*, vol. 4, 1926, p. 91.

[4] D. B. Mair, "The Valuing of English Composition", *Journ. of Educ.*, November 1928.

added to each letter, and the use of the Bryan Sheet to ensure normality of distribution.

Another writer[1] favours the analytic method where separate marks are given for Content, Structure and Mechanics. The Content has two and a third times the weighting of Structure, and the Structure has twice the weighting of Mechanics. The criticism that this would take time is met by insisting that it is more time that should be devoted to composition.

A more elaborate use of the analytic method is that of Wallis,[2] who marks seven points, namely: vocabulary, accuracy, craftsmanship, consistency, completeness, quantity of ideas, quality of ideas. A mark of 1, 3, 5 or 7 is awarded for each of these points, which means that the steps of difference jump by 4 per cent. The first five points cover technical skill and are marked irrespective of the subject set.

After considerable experience as Chief Examiner, Wallis is of opinion that this scheme is reliable for marking essays of young children in scholarship examinations. It causes the average candidate to score approximately half marks, it spreads the marks in the fashion which the modern examiner desires, and its results correlate closely with total marks for the whole Examination.

WHAT TRAITS ARE MEASURABLE?—It is usual to measure physical traits such as height or strength and mental traits such as technical knowledge or intelligence. And for the moment it is assumed that the probable errors of such measurements are known. On the other hand, some speak of traits like honesty which an individual is deemed to possess, or not to possess; that is, they assume that an individual must be honest, or else dishonest. Further, some speak of traits such as sociability or tact which may be judged but not measured.

In the last analysis it may be possible to dispense with such distinctions. For if it is possible to rate an individual relative to his fellows with respect to his honesty or his tact, there seems no reason why his honesty or his tact should not be

[1] G. Perrie Williams, "The Marking and Standardising of Composition", *ibid.*

[2] B. C. Wallis, *The Technique of Examining Children*, 1930, p. 77.

regarded as measurable. Let five competent judges who are acquainted with twenty individuals rate them on a scale where the maximum is ten, both with respect to their tact and also with respect to their intelligence. Each person then receives an average mark for each of the two traits. By comparison with these averages, the average deviation of each judge in assessing tact may be compared with his average deviation in assessing intelligence, and the relative difficulty of marking each may be ascertained. Such a method then may be applied not only to products such as drawings or essays, but to traits such as sociability or tact. Those who would limit the use of measurement should bear these possibilities in mind, they should also realise that there is no measurement without error, not even the simplest measurements of length. Thus Myers[1] refers to two competent observers who measured the head of a pygmy. One found the length and breadth to be 198 mm. and 147 mm., which yielded a cephalic index of 74·24. For the other the corresponding figures were 195 mm., 150 mm. and 76·92.

Hollingworth[2] gives the results of experiments on such traits as neatness, intelligence, humour, conceit, beauty, vulgarity, snobbishness, refinement and sociability :

(1) An individual's error in judging himself is somewhat greater than the average error of friends.

(2) There is a tendency for an individual to overestimate himself on desirable traits and to underestimate himself on the undesirable.

(3) In the traits of neatness, intelligence, humour, refinement or sociability, the individual who ranks high in the trait in the estimate of his associates tends to be a better judge of that trait in others than the individual in whom the trait is less conspicuous.

THE SELECTION OF APPLICANTS.—Whenever an applicant is being judged as to his suitability for a post the process of marking is more or less explicitly in operation. Thus some or all of the following seven devices may be employed: (1) The Letter of

[1] C. S. Myers, "The Pitfalls of 'Mental Tests'", *Brit. Medic. Journ.*, January 28, 1911.

[2] H. L. Hollingworth, *Judging Human Character*, 1922, chap. 4.

Application. (2) The Photograph. (3) The Interview. (4) The Testimonial. (5) The Rating Scale. (6) The Scholastic Record. (7) Psychological Tests. It is necessary to study each device and to ascertain what validity, if any, attaches to the marking or judging as it is usually carried out in practice.

(1) *The Letter of Application.*—Probably all would agree that the letters of some applicants might at once debar them from further consideration, but that is far from saying that an inspection of a batch of letters would enable a judge to choose the most suitable applicant. Hollingworth [1] found that judges disagreed with each other when instructed to arrange twenty-five letters in order of merit with respect to intelligence, reliability, tact or neatness. Further, a particular judge may disagree greatly with himself if he judges them again a month later. Not much value, then, should be attached to judgments based on letters of application. In order to improve the technique, Hollingworth suggests that the capacity of the judges for their task should be checked by noting the subsequent success of the applicants. Another improvement is to employ the Application Blank where the candidate fills in certain relevant data whose relative importance has already been decided.

(2) *The Photograph.*—Here again it would appear that experiments by Pintner [2] and others have shown that the judgments of mental traits from photographs are practically valueless. Thus it is practically impossible to judge " intelligence" from a photograph.

(3) *The Interview.*—In its traditional form the interview is highly unreliable. Hollingworth [3] found that twelve expert sales managers were quite unable to agree in their judgments on 57 applicants for appointment as salesmen. A candidate was likely to be ranked anywhere from highest to lowest by his twelve judges.

Magson [4] made a thorough analysis of the estimates made by

[1] H. L. Hollingworth, *op. cit.* chap. 2.

[2] R. Pintner, "Intelligence as Estimated from Photographs", *Psychol. Rev.* vol. 25, 1918.

[3] H. L. Hollingworth, *op. cit.* chap. 5.

[4] E. H. Magson, "How we judge Intelligence", *Brit. Journ. of Psychol.* Mon. Suppl. No. 9, 1926.

interviewers and concluded that the interview is practically worthless as a means of measuring general ability. The conversation covers too small a field and the judgments are made upon fallacious indications such as the manner, facial expression and personal appearance of the subject. Particularly important is his conclusion that when the interview is employed for selecting children for free places in secondary schools, it is essential that the interviews should be in charge of skilled interviewers under rigidly controlled conditions.

Recently attempts have been made to improve the validity of the interview. Spielman and Burt[1] selected character qualities, both elementary and complex, and developed a rating scale for their assessment. Two psychologically trained investigators independently interviewed thirty children with special reference to these qualities. For some qualities their impressions correlated highly: submissiveness 0·85, self-confidence 0·77, fear 0·75, assertiveness 0·74, sociability 0·72, anger 0·71. For others the correlations were distinctly lower: acquisitiveness 0·23, reliability 0·36, curiosity 0·37, punctuality 0·44, honesty 0·52, industry 0·54. Thus temperamental qualities which might naturally arise at the interview were found easiest to estimate. On the other hand, where the quality to be assessed is moral rather than temperamental, then its estimate cannot be relied upon, for its expression may be successfully concealed.

It was found, moreover, that practice improves the power of interviewing.

O'Rourke[2] has developed a scheme for training interviewers. The possible answers to the questions considered essential for the interview are printed and assessed on a form. After training and practice with this form it was found that the ratings of the interviewers improved in validity.

(4) *The Testimonial.*—The mention of references and testimonials sometimes raises a smile. It is true that if individuals were judged solely from testimonials the judgments would

[1] W. Spielman and C. Burt, "The Estimation of Character Qualities in Vocational Guidance" in Report No. 33, Indust. Fatigue Research Board, 1926, p. 57.

[2] L. J. O'Rourke, *A New Emphasis on Federal Personal Research and Administration*, 1930.

hardly furnish a normal curve! It would probably be a skewed curve minus its tail! Nevertheless, unless all faith in human nature is lost, it must be admitted that references and testimonials may be useful not only for what they say but also for what is left unsaid. There are many factors which affect their validity. If sent to strangers they may be less carefully prepared than when sent to acquaintances. In some cases references as to ability are more reliable than those concerning character as they are based on objective evidence, and as there is a tendency to rate character high in the absence of evidence to the contrary. In other cases, testimonials as to character based on long acquaintance by judges of repute may lack nothing in reliability. Testimonials are often misleading owing to a former employer being too kind-hearted to give his honest opinion of an individual even if his services have been unsatisfactory. But on the other hand an employer, jealous of an individual's progress, may attempt to damn him with faint praise, sometimes unsuccessfully and with amusing consequences.

(5) *Rating Scales.*—A rating scale dividing individuals into "good" and "bad", or even into "good", "bad" and "indifferent", with respect to any trait or ability, would be too coarse to prove very efficient.[1] A division into five classes such as excellent, good, average, poor, very poor would be better, but it is difficult to ascertain if one rater's assessment of "good" means the same as another's. The difficulty remains when the figures 5, 4, 3, 2, 1 are substituted for the adjectives. In consequence attempts have been made to use more specific terms referring to the individual's actual behaviour with reference to the ability which is rated, and also to use specific definitions for the ability to be assessed and not rely on single words which may be highly ambiguous.

A further development, especially when it is desired to secure

[1] Bartlett, in his investigation of the value of advertisements, sorted them into "good", "medium" and "poor" and then used his "method of fractionation", namely, to sort each group into three sub-groups and then combine bottom "good" with top "medium", and bottom "medium" with top "poor". He thus obtained seven groups the sizes of which approximated to normal distribution. R. J. Bartlett, "The Judgment of the Value of Advertisements and the Construction of Rating Scales", *Journ. of Nat. Inst. of Indust. Psychol.* vol. 3, 1927.

separate ratings on a variety of aspects of a vocation is the Graphic Rating Scale. The trait which is assessed is represented by a straight line with various descriptive expressions underneath. The scorer places a check on the line at the point which in his estimation corresponds to the amount of that trait possessed by the person who is rated, *e.g.*:

QUALITY	REPORT				
Success in winning respect and good-will through his personality		√			
	Inspiring	Favourable	Indifferent	Unfavourable	Repellent

The success of a rating scale would seem to depend on the skill and training of the rater, always assuming that the scale has been drawn up by an expert so as to ensure that no ambiguous terms occur in it and that no personality traits too difficult for assessment are included.

(6) *The Scholastic Record.*—Valentine[1] has recently carried out a comprehensive inquiry into the reliability of examinations for estimating the future academic success of candidates. Especially startling and disquieting are his conclusions on the prognostic reliability of entrance examinations to secondary schools. There was practically no relation between the order of merit in the entrance examination and the order of merit at the end of the secondary school career. Valentine concludes tentatively that the chief causes for this discrepancy are:

(*a*) The special cramming of certain pupils.

(*b*) The superior teaching in certain schools as compared with others.

(*c*) The inability of an examination in arithmetic and English (and possibly even of good tests of general intelligence) to indicate relative ability in the varied subjects studied in the secondary school.

(*d*) Different rates of mental development and the gradual influence of good and bad qualities of character.

(*e*) The "instability" of the entrance examination in the

[1] C. W. Valentine, *The Reliability of Examinations*, 1932.

sense that a similar examination a few days later might give very different results, at least, within a considerable range of the critical border-line.

Valentine's suggestions for improvement deserve close attention: The extension of the use not only of the best type of intelligence tests but also of tests of specific abilities needed in the secondary school; the better standardisation of the reports of the heads of elementary schools; the providing of a "second chance"; the transition to a secondary type of education at a later age than eleven plus; and a reconsideration of awards in the secondary school at the end of the first year.

It thus becomes an urgent problem how best to avoid the great loss due to the fact that some pupils deserving of scholarships do not receive them.

As Sandon[1] has pointed out, prediction is a question of probabilities, and selection modifies means, standard deviations and correlations. Thus the correlation between test and criterion is almost worthless unless the degree of selection is known.

Barraclough's researches[2] have shown the importance of testing the significance of any correlations between entrance and school certificate examinations. By the methods of analysis of variance Barraclough found that out of a total of 26 investigations regarding the predictive value of English at entrance, 16 scholarship examinations were non-predictive, while 10 had predictive values ranging between zero and 32 per cent. Arithmetic was only predictive of Mathematics in 7 cases out of 27, the maximum predictive value being 40 per cent.

Barraclough refers to the microscopic attention often accorded to "border-line cases". Yet there can be no confidence that these children really are border-line for most scholarship examinations have a common defect—lack of standardisation.

Further the important investigation of Wilson[3] shows that

[1] F. Sandon, "Difficulties in using Entrance Examinations, Intelligence Tests and School Results for Comparative Purposes", Rep. of Brit. Assoc., Section L, 1933.

[2] F. Barraclough, "The Reliability of Entrance Examinations to Secondary Schools", *ibid.* 1933.

[3] J. H. Wilson, "Group factors among Abilities involved in a School Certificate Examination", *Brit. Journ. of Educ. Psychol.* vol. 3, 1933, pp. 71-86, 99-108.

secondary school examinations measure neither the pupils' general capacity nor their specific abilities at all adequately and therefore the discovery of efficient tests of special abilities as well as tests of general capacity is of the utmost importance.

Investigations such as those of Stead,[1] and especially those of Oates,[2] are promising in that they have brought to light the bearing of ratings and even of tests of temperament on scholastic achievement.

The important statistical papers of Sandon throw light on many problems of marking. Some of Sandon's conclusions may thus be summarised: (1) If two or more divisions of a school form have been selected by random sampling, it cannot be assumed that a pupil who gets a certain total percentage of marks in one division has the same scholastic ability as one who gets the same percentage in another division. (2) In the case of school subjects which may be recognised as of equal importance, it does not follow that these subjects are weighted equally if the marks in each subject are brought to the same maximum, but they can be weighted equally enough for practical purposes by making the quartiles in each case the same. Sandon's method was to put the quartiles equal to 60 and 40 and then scale the marks by means of a nomogram. (3) The age allowance in entrance scholarship examinations to secondary schools should be a percentage of maximum marks and not of actual marks.

Probably most would agree that age allowances are absolutely necessary at eleven plus, and most would also be in agreement with Godfrey Thomson: "Personally I prefer that no under age candidates be admitted at all, which is I think both socially and statistically sound. Eleven is young enough, and a second chance can be given at twelve." It may be left to the reader to draw his own conclusions as to the best method of awarding age allowances. The following references will be

[1] H. G. Stead, "Factors in Mental and Scholastic Ability", *Brit. Journ. of Psychol.* vol. 16, 1926.

[2] D. W. Oates, "(1) An Experimental Study of Temperament", *ibid.* vol. 19, 1928; "(2) Group Factors in Temperament Qualities", *ibid.* vol. 20, 1929; "(3) The Relation of Temperament and Intelligence to Scholastic Ability", *Forum of Education*, vol. 7, 1929.

found useful: (1) E. J. G. Bradford, "Selection by Examination", *Forum of Education*, vol. 1, 1923 ; (2) F. Sandon, "The Scaling and Totalling of School Marks", *ibid.* vol. 2, 1924; (3) F. Sandon, "A Statistical Analysis of Some School Marks", *ibid.* vol. 3, 1925; (4) A. James, "The Age Allowance of Marks in Secondary School Admission Examinations", *ibid.* vol. 3, 1925; (5) F. Sandon, "A Statistical Study of Some School Marks: The Fallacy of Percentages", *ibid.* vol. 4, 1926; (6) A. Bell, *Report of an Investigation of the Free Place Scholarship Examination in the County of Kent*, 1926; (7) J. W. Collier, "The Treatment of Secondary Schools' Admission Examination Scores", *ibid.* vol. 6, 1928; (8) *Free Place Examinations*, Bd. of Educ. Educational Pamphlet No. 63: Memorandum on Examinations for Schools and Free Places in Secondary Schools, 1928; (9) F. Sandon, "Some Effects of Age in Selective Examinations," *Forum of Education*, vol. 6, 1928 and vol. 7, 1929; (10) B. C. Wallis, "The Efficiency of Competitive Scholarship Examinations of Young Children with particular reference to the Effect of an Age Allowance", *ibid.* vol. 8, 1930; (11) *Examinations in Public Elementary Schools*, Report of the Joint Comm. of Associations of Educ. Committees and the Nat. Union of Teachers, 1930 (cf. p. 200); (12) F. Sandon, "The Basis of Marking", *Brit. Journ. of Educ. Psychol.* vol. 1, 1931; (13) G. H. Thomson, "The Standardisation of Group Tests and the Scatter of Intelligence Quotients: A Contribution to the Theory of Examining", *Brit. Journ. of Educ. Psychol.* vol. 2, Parts 1 and 2, 1932 (especially p. 103 and p. 132); (14) C. W. Valentine, *The Reliability of Examinations*, 1932 (cf. p. 182); (15) Memorandum on Age Allowance by the Bradford Educ. Comm. 1933; (16) F. Sandon, "Progress through the Secondary School as measured by School Marks", *Brit. Journ. of Educ. Psychol.* vol. 3, 1933.

Sufficient has been said to show that considerable attention has been devoted to improve the reliability and validity of marking. It is quite clear that marking is an extremely difficult task and that considerable injustice is often done. There is urgent need of further research in many directions. There are, further, many theoretical difficulties: On what basis or bases

should (*a*) one subject be regarded as important as another or (*b*) one question be regarded as important as another? Should the criterion be pure impressionism, or should it be based on relative difficulty, or on the time which each demands, or on the correlation of each with "g", or with some other factor? Who knows?

(7) *Psychological Tests.*—Cf. Chapter 8.

QUESTIONS

1. What method did you adopt in marking the seven essays? What difficulties were experienced?
2. Essays have been appraised by comparing them with a standard scale of samples. What are the advantages and disadvantages of the method?
3. "Capable of great labours—as will be seen—capable of great self-sacrifice, steady in conviction, even amid encompassing gloom and disheartening isolation, he was wanting in the activity, the tenacity, the ceaseless industry, the changeless purpose, and also in the rough and almost brutal self-assertion which is one of the conditions of success in most walks of life" (T. P. O'CONNOR's memoir of Lord Morley, *Daily Telegraph*, September 24, 1923). Consider the possibility of rating the traits mentioned.
4. Show in detail how you would statistically deduce the relative importance of the four factors mentioned in the text which affect school marks.
5. Classify the mistakes made by school children in simple addition, and illustrate the difficulty of marking even in arithmetic. (T. N. K. Rao, "An Experimental Study into the Backwardness of Elementary Schoolboys in Arithmetic", M.Ed. Thesis, Univ. of Leeds, 1934).

REFERENCES

1. F. Y. Edgeworth, "The Statistics of Examinations", *Journ. of Royal Stat. Soc.* vol. 51, 1888; pp. 599-635; also "The Element of Chance in Competitive Examinations", *ibid.* vol. 53, 1890, pp. 460-475, 644-663; also "Problems in Probabilities", *Phil. Mag.* August 1890.

2. P. J. Hartog, *Examinations and their Relation to Culture and Efficiency*, 1918.
3. D. Starch, *Educational Psychology*, chap. 22, 1920.
4. H. L. Hollingworth, *Judging Human Character*, 1922.
5. P. B. Ballard, *The New Examiner*, 1923.
6. W. Boyd, *Measuring Devices in Composition, Spelling and Arithmetic*, 1924.
7. J. M. Crofts and D. Caradog Jones, *Secondary School Examination Statistics*, 1928.
8. C. W. Valentine, *The Reliability of Examinations*, 1932.

CHAPTER 21

THE DETECTION OF SUPPRESSED IDEAS

MATERIALS.—Psycho-galvanic reflex apparatus, stop-watch, lists of words.

DIRECTIONS.—Waller's apparatus, made by Messrs. Gambrell Bros., Ltd., London, may be used for this experiment. It is essentially a Wheatstone bridge in one arm of which S is placed. The electrodes are zinc discs covered by chamois leather and moistened with 6 per cent saline solution, and are applied to the palm and dorsum of the left hand, and kept in position by indiarubber bands. There is a shunt to reduce the sensitivity of the galvanometer when the preliminary adjustments of the bridge are being carried out.

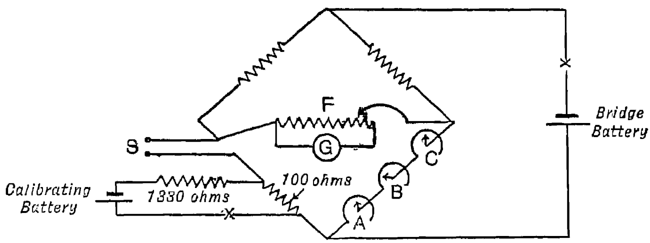

FIG. 26.

The bridge current is derived from two dry cells and can be switched on or off as desired. In series with S is a resistance of 100 ohms which is connected with a resistance of 1330 ohms and one dry cell which is governed by another switch called the calibrating switch, and hence a current of one milliampere can be passed around the circuit. In making an experiment the following procedure is adopted: The galvanometer is permanently set up on a fixed support attached to the wall, and the lamp and scale are set up at one metre distance, and the lamp focussed until a clear spot is obtained on the scale. The galvano-

meter is connected to the Galvo. terminals, S to the terminals marked "Patient", with the left hand resting comfortably on the table, and the switches and the shunt are set at "Off". Everything being ready the battery switch is switched on, the shunt is turned to the mark 500 and the galvanometer deflects slightly and is brought to zero by adjustment of the rheostats. This rotation of the shunt with the corresponding adjustment is made until finally the shunt is set at the mark 10, 5 or even 2, according to the size of the reflex. In other words the deflections obtained should neither be too small nor too large for measurement. Before taking a record the calibrating switch is switched on for about three or four seconds to allow the galvanometer to attain maximum deflection and is then switched off. This gives the calibrating deflection. A stimulus is then applied and the response is indicated by the movement of the beam of light. Should the response deflection be too large it can be reduced by adjustment of the shunt, but the calibrating must be done in the same shunt position as that on which the response was taken. For the purpose of the present experiment the ratio of the response deflection to the calibrating deflection may be called the "psychogalvanic ratio".

For most Ss there is a tendency for the resistance to decrease gradually for the first few minutes and so it may be necessary to wait a few minutes for the spot to become steady.

The response deflection has a latent period, usually between 2 and 3 seconds, and thus the spot of light does not immediately move after the stimulus is applied. When necessary a screen may be used to shut off the apparatus from S's view.

The experiment now to be performed illustrates the "third degree" method of detecting crime. Two subjects are selected and having received a sealed envelope, they leave the room together and decide which is to open the envelope and commit the "crime". The other must know nothing about the contents of this sealed envelope or of the "crime". The sealed envelope contains directions such as the following: "Go into my room and open the letter lying on my desk. Read it carefully. When you are examined afterwards try as well as you can to conceal the fact that you have seen the letter."

It is not easy to introduce into the laboratory a situation where S is emotionally affected as would probably be the criminal subjected to the "third degree" examination, but it is possible to demonstrate the procedure even when the contents of the letter are hardly conducive to the stirring up of emotion. For this purpose the following statement about "Sanjam" may serve as a brief illustration. It is sealed in an envelope and placed in a box containing an ornamental bird tallying with the description given.

Sanjam

Sanjam is the cruel god of the heathen tribe San in Central Africa. Its beak is red with the blood of sacrifice and its talons yellow with the gold which is daily offered to appease its wrath. Its body is tinged with the blue of the sacred river into which its helpless victims are thrown. More than fifty thousand human lives have been sacrificed to Sanjam.

When the "criminal" has completed operations and has returned to the "innocent", E is informed by them that the instructions have been carried out. E now calls in one of the two Ss for examination, the members of the class having been in the meantime informed of the details of the "crime". E has arranged a series of 30 words, of which 12 are related to the "crime" and are called "crucial words", the remaining 18 being "non-crucial". It is better to commence with several non-crucial words. The words in the series were: hat, dinner, window, telephone, pencil, *god*, *Sanjam*, tree, *beak*, *sacred*, church, coffee, *cruel*, star, *fifty*, *central*, *talons*, storm, orange, *blue*, fire, John, class, *heathen*, ancient, chalk, *sacrifice*, *river*, history, brick.

When S's left hand has been joined to the terminals and his resistance measured, E speaks the first word of the series, having previously asked S to reply as quickly as possible with the first word that occurs to him. E notes on the stop-watch the time in fifths of a second that elapses between the giving of the stimulus-word and the reaction of S, while an assistant manipulates the galvanometric apparatus and notes the deflection

corresponding to each reaction. The results may be tabulated thus:

SUBJECT X

Stimulus-word	Reaction-word	Reaction-time	Galv. Deflection

After the experiment is completed the same procedure is carried out with the other S. Then the average reaction-time for the crucial words and also for the non-crucial words is calculated for each S, then the difference between the average crucial and the average non-crucial reaction-times for each S is expressed as a per cent of the average of each S's non-crucial reaction-times. This is an important measure and is called the "crucial difference". In precisely the same manner the crucial difference with respect to the galvanometric deflection may be calculated. In deciding which of the two Ss is "guilty" all the following points are to be considered as possible evidences of guilt.[1]

(1) Greater crucial difference in galvanometric deflections.[2]

(2) Unusually large deflections for certain crucial stimulus-words.

(3) Greater crucial difference in reaction-times.

(4) Greater mean variation in reaction-time to the crucial than to the non-crucial words.

(5) Unusually long reaction-times for certain crucial stimulus-words.

(6) Reaction-words connected with the "crime".

(7) Reaction-words showing unusual associations.

(8) Failure to give a reaction-word.

(9) Evidence of a plan of reaction arranged by S beforehand.

To hide all traces of "guilt" would require much cleverness on S's part. There are, however, two or three devices which a student well conversant with recent literature might attempt.

DISCUSSION

The foregoing experiment is obviously a combination of two distinct experiments, namely, the association experiment and

[1] Cf. H. S. Langfeld and F. H. Allport, *An Elementary Laboratory Course in Psychology*, 1916, p. 115.

[2] In actual practice this was not found easy to apply as many deflections were too slight.

the *p.g.* reflex experiment. Taken simultaneously they serve as an interesting demonstration, and students are able to judge how far the indications of each experiment correspond. There is no unanimity of opinion as to the physiological causation of the *p.g.* reflex. So many respiratory, circulatory, muscular and secretory changes, either singly or collectively, have been suggested as explanatory of the reflex that no useful purpose would be served by a detailed discussion here.

There is even no agreement concerning the psychological significance of the reflex. Although it has been customary to regard the phenomenon as an indication of emotion, and to call by the name "emotometer" the apparatus for measuring the reflex, yet another view has been advanced by Aveling[1] and his pupils. As a result of recent experiments they were led to the view that the reflex is characteristically the consequent of conation, and that consciously experienced "emotion" is not the cause of it. The term "emotion" is often loosely used for "affection" (pleasure-unpleasure), but Aveling understands by emotion the "stirred-up" character of consciousness due to somatic resonance, and as this is most frequently subsequent to the mental process which terminates in the reaction, it cannot be the cause of the reaction. By an ingenious technique introspections with reference to will-acts (resolves) as well as to conation (trying or striving) were obtained. It was found that deflections occur when conation is experienced, and that they do not always occur when affection or emotion is experienced. Moreover, the deflections were greater for conative than when affective experience is reported, and varied in size roughly with the estimated intensity of the conative experience. It was also found that deflections occurred in the absence of affective experience; and, the deflections are greater when conative experience is reported than when will-acts take place.

R. J. Bartlett, however, argues that the mental basis of the phenomenon is a *complex* orectic process rather than "conation" or "emotion" alone, and that as a state of passive endurance

[1] F. Aveling, "The Conative Indications of the Psychogalvanic Phenomenon", *Proc. of VIIIth Internat. Congress of Psychol.*, Groningen, 1926, p. 227.

or enjoyment may pass into a more active one, the search for a crucial test of the mental cause is difficult.[1]

Mention may here be made of some preliminary tests which I undertook at a time when the prevailing opinion pointed to emotion as the causative factor. In the first series of preliminary tests, the subject assumed four different attitudes at different sittings: (1) he was engaged in reading a book when the stimuli were applied; (2) he concentrated on the stimulus with a determination not to be affected when it occurred; (3) he concentrated on keeping the limbs and trunk equipoise and all his muscles lax with a view to resist being affected emotionally; (4) he concentrated on specific problems quite unconnected with the experiment, *e.g.* made a list of the health resorts in the North of England. In (2), (3) and (4), at some point within 10 seconds after the attitude had been taken, the stimulus was applied; especially noticeable was the fact that taking up any of these three attitudes caused a deflection which might be greater, equal or less than the subsequent deflection on application of the stimulus. The latter deflection, however, was less than the deflection in (1).

Another series of tests was commenced with a group of boys, some stolid and some emotional according to their teacher's report. Three sets of deflections were obtained: (1) when the stimuli were applied without warning; (2) those due to a warning "Now"; and (3) those which followed the stimulus when this was applied shortly after the warning "Now". By this means it was intended to ascertain which set of deflections, if any, correlated with the teacher's report.

MUHIYUDDIN'S EXPERIMENTS WITH SCHOOL CHILDREN.—In 1921 Muhiyuddin[2] studied the *p.g.* reflex with Leeds children. His subjects were 47 boys ranging in age from eleven years eleven months to twelve years eleven months, and 10 University students between twenty-one and twenty-five years.

Influenced by the introspective reports of his own subjects and also by the evidence of Waller, Starch, Binswanger,

[1] R. J. Bartlett, "Does the Psychogalvanic Phenomenon indicate Emotion?" *Brit. Journ. of Psychol.* vol. 18, 1927.

[2] M. S. Muhiyuddin, "An Experimental Study of Emotion", M.Ed. Thesis, Univ. of Leeds, 1922.

Peterson and Jung, and others, Muhiyuddin accepted the view as to the emotional causation of the reflex, but some of his results are unaffected by that controversy. Thus as each S was tested twice, once in the morning and once in the afternoon, it was possible to get some date as to the diurnal variation in resistance, the electrodes being applied to the palm and dorsum of the left hand. It was found to be higher generally in the morning (9·45-12) than in the afternoon (1·30-4), the medium values being 10,625 ohms and 7312 ohms respectively. Experienced teachers were asked to award to each pupil the marks +2, +1, 0, −1, or −2 in respect of each of the following qualities:

(1) Readiness to feel (*a*) fear, (*b*) anger.
(2) Intensity with which he feels (*a*) fear, (*b*) anger.
(3) Duration of the state of (*a*) fear, (*b*) anger.

It was found that the marks of one teacher agreed but little with those of another although both were well acquainted with the pupils, and that there was no correlation between the size of the reflex and the estimates of emotivity furnished by the teachers.

The same teachers showed very fair agreement in assessing the general intelligence of their pupils, the coefficient of correlation being 0·567, which compares favourably with figures given by other investigators. Yet the essentially inner character of the emotions escape the searching eye of an observant teacher. A boy may even be classed as "nervous" by one teacher and "stolid" by another.

Muhiyuddin also found a correlation of 0·43 (P.E. 0·06) between "intelligence" as measured by tests and *p.g.* reflex when produced by "ideational" stimuli or "threats"; also a correlation of 0·26 between "intelligence" as measured by tests and *p.g.* reflex when produced by "sensory" stimuli, *e.g.* the report of a toy pistol.

Miss M. D. Waller[1] also concluded that the size of the *p.g.* reflex is in some degree associated with intellectual efficiency, but this was assessed only from the result of an hour's ex-

[1] M. D. Waller, "The Emotive Response of a Class of 73 Students of Medicine", *Lancet*, 1918, p. 510.

amination in physics, the subjects being women students of medicine.

Prideaux[1] also as a result of his experience formed the view that "the greater the intellectual development the more pronounced are the reactions to ideational stimuli."

W. S. Brown[2] found the size of the reflex to correlate as follows with teachers' estimates of intellectual qualities: Intelligence, 0·09. Quickness of apprehension, 0·20. Profoundness of apprehension, 0·12. Soundness of common sense, 0·22. These coefficients are very low, as the probable error is 0·09. Nevertheless it is too early to come to a definite conclusion, for as Thouless[3] points out: "The problem of what is a significant measure of the P.G.R. is not yet solved". . , . "We call out the word 'dog' and our subject's apparent resistance changes from 10,000 ohms to 9000 ohms. Ten minutes later, when we come to the word 'cat', his apparent resistance is 7000 ohms. What change of resistance would indicate that his reaction to 'cat' was equal to that to 'dog'?" This difficulty is especially acute when different individuals varying widely in apparent resistance are compared and ranked with respect to the size of their P.G.R.'s. In this chapter I have largely avoided that difficulty by choosing experiments where different stimuli or different instructions are employed with the *same* individual.

QUESTIONS

1. Enumerate some of the difficulties experienced when attempts are made to rank subjects according to the size of their psychogalvanic responses.
2. What importance would you attach to the P.G.R. Test if employed in a court of law as a means of detecting deception?

[1] E. Prideaux, "The Psychogalvanic Reflex: A Review", *Brain*, Part 1, 1920, p. 50.

[2] W. S. Brown, "A Note on the Psychogalvanic Reflex considered in Conjunction with Estimates of Character Qualities", *Brit. Journ. of Psychol.* vol. 16, Part 2, 1925.

[3] R. H. Thouless, "The Technique of Experimentation on the Psychogalvanic Reflex Phenomenon and the Phenomenon of Tarchanoff", *Brit. Journ. of Psychol.* vol. 20, Part 3, 1930, p. 230.

References

(Some recent references, additional to those given in the text)

1. H. M. Wells, "A Note on the Psychological Significance of the Psycho-galvanic Reflex", *Brit. Journ. of Psychol.* vol. 14, 1924.
2. E. Farmer and E. G. Chambers, "Concerning the Use of the Psycho-galvanic Reflex in Psychological Experiments", *ibid.* vol. 15, 1925.
3. R. H. Thouless, "The Causes of the Continuous Change of Resistance observed in Psycho-galvanic Experiments", *ibid.* vol. 16, 1925.
4. D. Wechsler, "The Measurement of Emotional Reactions: Researches on the Psycho-galvanic Reflex", *Arch. of Psychol.* No. 76, 1925.
5. H. E. O. James and R. H. Thouless, "A Note on the Effect of Polarisation in Psycho-galvanic Reflex Experiments", *Brit. Journ. of Psychol.* vol. 17, 1926.
6. H. C. Syz, "Observations on the Unreliability of Subjective Reports of Emotional Reactions", *ibid.* vol. 17, 1926.
7. H. C. Syz, "Psycho-galvanic Studies on Sixty-four Medical Students", *ibid.* vol. 17, 1926.
8. C. W. Darrow, "Sensory, Secretory and Electrical Changes in Skin following Bodily Excitation", *Journ. of Exp. Psychol.* vol. 10, 1927.
9. E. G. Flemming, "Personality as revealed by the Galvanometer", *Amer. Journ. of Psychol.* vol. 38, 1927.
10. R. B. Cattell, "The Significance of the Actual Resistances in Psycho-galvanic Experiments", *Brit. Journ. of Psychol.* vol. 19, 1928.
11. L. A. Jeffress, "Galvanic Phenomena of the Skin", *Journ. of Exp. Psychol.* vol. 11, 1928.
12. R. B. Cattell, "Experiments on the Psychical Correlate of the Psycho-galvanic Reflex", *Brit. Journ. of Psychol.* vol. 19, 1929.
13. C. Landis and H. N. De Wick, *The Electrical Phenomena of the Skin (Psycho-galvanic Reflex)*, Psychol. Bull. No. 26, 1929.
14. C. A. Ruckmick, "Emotions in Terms of the Galvanometric Technique", *Brit. Journ. of Psychol.* vol. 21, 1930.

CHAPTER 22

ILLUSIONS

THE MÜLLER-LYER ILLUSION

Experiment 1

MATERIALS.—Black lines are drawn on thin xylonite as shown in the diagram. The short lines at the ends form an angle of 120 degrees. The line on the left is 100 millimetres long. The line on the right is variable and is drawn on a part which slides in and out of a framework directly under the other part. At the back is drawn a scale in millimetres which enables E directly to read how many millimetres shorter or longer than the constant line is the variable when they are presented to S, or when S adjusts the variable so that it appears equal to the constant.

In case xylonite is not available, white cardboard may be used. By suitable illumination the vertical line dividing the two parts may be caused to disappear if S looks at the figure slightly obliquely from a distance of about 18 inches.

PROCEDURE.—Two methods may be employed.

(*a*) S himself adjusts the variable to appear equal to the standard. Five readings may be taken, starting from each of four positions, namely, with the variable on the right and distinctly longer than the standard, with the variable on the right and distinctly shorter than the standard, together with the two corresponding positions when the variable is on the left. The Crude Constant and Crude Average Errors and the Space Error should be calculated (cf. Chapter 24).

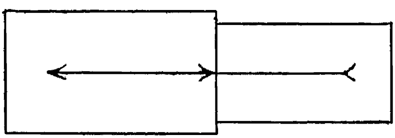

FIG. 27.

(*b*) *The Constant Method.*—E sets the variable at definite positions, according to the millimetre scale. The figure is then

exposed for two seconds and S judges whether the variable is longer, shorter or equal to the standard. Here again the variable should appear on the left as well as on the right, and several exposures for each position should be given (cf. Chapter 24).

The results obtained by the two methods should be compared.

THE SIZE-WEIGHT ILLUSION

Experiment 2

MATERIALS.—Two standard boxes and eighteen comparison boxes. Pill boxes form a suitable material. Both standards weigh 55 grams; both are 28 mm. thick, but the larger is 82 and the smaller 22 mm. in diameter. The eighteen comparison boxes are all 28 mm. thick and 35 mm. in diameter, but their weights range from 15 to 100 grams by 5 gram increments.

PROCEDURE.—Arrange the eighteen comparison boxes on thick felt on a table, in the order of their weight from left to right. Place before S the larger standard, and say: "Here is a box. I want you to find a box in this series of eighteen boxes that seems to you just as heavy as this one. Lift it by picking it up edgewise with your thumb and finger. Then try the first of these weights (at the left). If that does not suit, try the next, then the third, and so on, till you find a box that seems equal to this one. Each time you must lift this block first, then the one you are trying in the series. Keep your eyes constantly directed at the weight you are lifting".[1]

When S has selected an equivalent weight, the same procedure is followed with the second, or smaller standard box.

PROGRESSIVE WEIGHTS

Experiment 3

MATERIALS.—A set of twelve weights, of identical size and appearance, numbered conspicuously from 1 to 12. The first four weigh 20, 40, 60 and 80 grams respectively; the remaining eight weigh 100 grams each.

[1] G. M. Whipple, *Manual of Mental and Physical Tests*, Part 2, Test 40, 1915.

PROCEDURE.—Place the twelve weights in a line as numbered, on thick felt on a table, with the lightest on the left and the eight heaviest on the right, and with about 2 cm. between each weight. No. 1 is then at the left, No. 12 at the right of the row. Say to S: "Here is a series of weights; twelve of them. I want you to lift them, one after the other, like this." (Illustrate by taking a weight between finger and thumb and lifting some 10 cm. from the table.) "As you lift each weight, I want you to tell me whether it is heavier, lighter, or the same as the one just before it. All you have to say is either 'heavier', or 'lighter', or 'the same'. Remember you are to compare each weight with the one you lifted just before. For instance when you lift the 8th, you are to say whether it is heavier, lighter, or the same as the 7th. Here is the first weight, number one, at the left end of the row."[1]

Watch S to see that he follows these instructions, particularly that he lifts the weights successively, without relifting earlier ones.

DISCUSSION

The study of illusion has throughout the ages presented many baffling problems. Since psychology became an experimental science considerable progress in the study can be recorded.

The publication of Sully's *Illusions—A Psychological Study* in 1881 helped to focus the attention of psychologists on the subject, and ever since they have been prolific in their explanations of the various illusions. But owing to the complexity of the subject it is still difficult to draw a sharp line between cases which are illusions and those which are not. One writer of fiction referred to "illusions" as the epithet applied by cynics to life's most beautiful realities. Such a description at least reminds the serious reader of the perennial philosophical conundrum—What is reality? To "the man in the street" illusion has a very loose meaning—a political heresy or a conjurer's trick, the delusions of a paranoiac or the hallucinations, whether entrancing or terrifying, of a drug addict. A student of psychology, on the other hand, would apply the term "delusion" to

[1] G. M. Whipple, *Manual of Mental and Physical Tests*, Part 2, Test 41.

a false belief which would entitle him to regard the individual holding it as abnormal; and McDougall[1] has classed all delusions into delusions of desire and delusions of aversion. But obviously no clear line can be drawn to indicate when a belief becomes a morbid delusion. Further the term "hallucination" is confined to misperceptions where sensory cues are absent, and although this distinction is practically convenient there are, strictly speaking, no cases of hallucination without some sensory cue. Thus no sharp line can be drawn between hallucination and illusion.

A narrower sweep than that of the "man in the street" but still a very broad sense is to confine the term "illusion" to a perception of a particular kind, namely, one which does not harmonise with objective reality as determined by the total available sensory evidence. Now illusion in that sense may be experienced owing to the conditions of stimuli outside the organism, *e.g.* mirrors, prisms or lenses, or it may be brought about owing to physiological processes inherent in the organism, *e.g.* retinal processes or processes of ocular muscles.

Examples of the first kind would be a succession of stationary pictures on a cinema screen producing an illusion of actual motion or the "ghost" produced on the stage by artificial lenses, or the mirage caused by the lenses of nature. Thus I once saw in Egypt a sheet of water which appeared as if it were situated in the desert although actually there was no water within many miles. Examples of the second kind would be the green after-image seen on a white sheet after fixating a red patch of colour, or seeing two fingers when one finger is held at a certain distance from the face. Confining our attention to the visual field, it is obvious that, in this broad sense, there would be manifold cases of illusion,[2] such as inverted images, after-images and after-effects of seen movement,[3] retinal rivalry, stereoscopic vision, parallax, convergence and accommodation, simultaneous and successive contrast, simultaneous and successive induction.

[1] W. McDougall, An Outline of Abnormal Psychology, 1926, p. 333.

[2] Cf. C. S. Myers, *Textbook of Experimental Psychology*, Part 1, chap. 6, and R. H. Wheeler, *The Science of Psychology*, 1929, chap. 3.

[3] Cf. A. Wohlgemuth, "On the After-effect of Seen Movement", *Brit. Journ. of Psychol.* Mon. Suppl. No. 1, 1911.

Usually, however, in psychological textbooks the term "illusion" is limited to the subjective perversion of the contents of objective perception, as in the case of the Müller-Lyer illusion (Experiment 1). Here the misinterpretation or perversion is surely different from that which obtains in the case of a mirage or of an after-image where psychological processes cannot produce the phenomena. The sensory cue is misinterpreted in the case of a mirage, but surely through no fault of the optical or cerebral systems. Now taking experience as signifying "something lived", it may be equally real in the case of the Müller-Lyer illusion, the mirage, and the after-image, and only when the phenomena are *absent* are doubts cast on the efficiency of the retina or cortex. Nevertheless, there is a distinct psychological process active in the illusion proper, the Müller-Lyer, which plays no part in the production of the mirage or the after-image.

Further, *all* individuals would experience such a phenomenon as a mirage, there would be no individual differences in susceptibility. In illusions proper, however, although they are called normal illusions, yet it is important to note that individuals do differ in their susceptibility to particular illusions. This could be strikingly illustrated by asking different persons to place a mark on the wall to indicate the height which a top hat would reach if placed on the floor. Or again, many of those who are fond of asserting the equivalence of a pound of lead and a pound of feathers would select lead weighing anything from a tenth of a pound up to half a pound as equal in weight to a pound of feathers. The word "normal" means nothing more than that most individuals are subject to the illusion to a greater or less extent. It may, however, be convenient in dealing with measures of an illusion to refer to the measure most commonly experienced as *normal*.

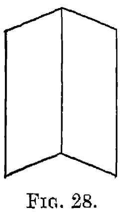

FIG. 28.

Confining our attention to optical illusions, they are usually classified under such headings as:

(1) *Illusions of Reversible Perspective.*—The figure may appear either as the inside or the outside of a book. Study the frequency of the changes. Analyse also the changes which occur when the

centre of the figure is fixated and its distance is increased or diminished.

(2) *Illusions of Reversible Figure and Ground.*—The figure can be seen as a row of black T's on a white background, or as a row of white leaves on a black background.

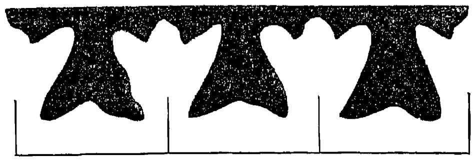

FIG. 29.

(3) *Illusions of Length.*—*E.g.* the Müller-Lyer illusion.

(4) *Illusions of Area.*—Cf. Question 1 at the end of this chapter.

(5) *Illusions of Direction.*—In Poggendorff's figure (FIG. 30) *cd* does not appear to be a prolongation of *ab*.

Experiment 1: *The Müller-Lyer Illusion*

This illusion has given rise to much discussion. Many eminent thinkers have advanced theories in explanation. Titchener[1] mentions twelve: (1) Delboeuf's Theory of "Attraction of Regard". (2) Brentano's Theory of the "Pseudoscopic Angle". (3) Auerbach's "Physiological" Theory. (4) Brunot's "Mean Distance" Theory. (5) Müller-Lyer's "Confluence" Theory. (6) Thiéry's "Perspective" Theory. (7) Wundt's "Eye-movement" Theory. (8) Einthoven's "Dispersion-Image" Theory. (9) Láska's Theory of "Joining the Discontinuous". (10) Heymans's "Movement Contrast" Theory. (11) Lipps's "Mechanical-aesthetic" Theory. (12) Jastrow's "Relativity" Theory.

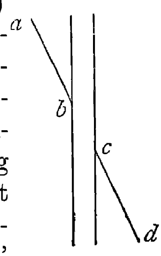

FIG. 30.

Those who wish to undertake a detailed study of this illusion will still find Titchener's account and list of original references indispensable.

[1] E. B. Titchener, *Experimental Psychology*, vol. 1, Part 2, 1901, p. 321.

Rivers[1] tested the degree in which English men, women and children, Todas, and Murray Islanders were subject to the Müller-Lyer illusion. The savage tended to be less subject to the illusion than the civilised man. As Myers[2] observes, the figure has less meaning to the savage and he is consequently less influenced by the figure as a whole when estimating the length of one of its parts. He is therefore able to attend more to that part and his task is more elementary.

Rivers also found that the extent of the illusion varied much less among his Papuan observers than among Englishmen, and similarly it varied less among English children than among English men.

The civilised adult is more acquainted with the illusion, and some may attempt by various means to overcome it. Thus his judgment is more complex and more variable than that of primitive man or civilised child.

Experiment 2: *The Size-Weight Illusion*

Of two objects of equal weight but varying greatly in size, the greater amount of muscular effort is put forth in lifting the bigger, and so it is judged to be the lighter of the two. Thus a pound of feathers appears lighter than a pound of lead. Normal adults are therefore subject to this illusion as a result of their experience, and it is not surprising to find that young or defective children are not subject to it. As a measure of the illusion one may take the difference in grams between the two blocks selected by S as the equivalents of the two standards.

The Material-Weight Illusion.—A solid cube of wood appears heavier than a hollow leaden cube of like size and weight, and if cork were substituted for wood the illusion would be stronger still. Here again there is a more or less unconscious inference which determines the amount of muscular effort put forth.

Experiment 3: *Progressive Weights*

This test is one of several devised by Binet in an attempt to measure the degree of suggestibility. As a measure of the

[1] W. H. R. Rivers, "Observations on the Senses of the Todas", *Brit. Journ. of Psychol.* vol. 1, 1905.

[2] C. S. Myers, *Introduction to Experimental Psychology*, 1911, p. 55.

illusion may be taken the number of times the judgment "heavier" is given in the last seven judgments which do not involve a real difference in weight between the stimuli.

Aveling and Hargreaves[1] tested children with a battery of Suggestibility tests which included the Progressive Weights test, and although they considered their evidence pointed to a general factor of Suggestibility complicated by group factors, there was no ascertained tendency for Suggestibility to go with other general factors such as General Intelligence.

ILLUSION AND GESTALT

Gestalt psychologists stress the fact that perceptual experience is never detached. It always has a setting. In other words, it is never dependent exclusively on its local stimulus. Köhler points out that "all the well-known 'illusions' may be cited as evidence of the fact that local processes depend upon *sets* of stimuli".[2]

Such a view commands general agreement. The real controversy centres round the fundamental doctrine of the Gestalt school, namely, that there are Wholes or Gestalten which are given *immediately*. Once this is granted, Gestalt-psychologists are able to account for illusions, at least *in general terms*. Thus Wheeler writes that the Müller-Lyer illusion "is conditioned by the total arrangement of the lines, various single factors contributing by virtue of their relationship to the whole".[3] Even so, probably Gestalt-psychologists would agree that there is the further need of some analysis of the "various single factors" which contribute to each particular illusion.

ILLUSION AND NOEGENESIS

Since the so-called "normal illusions" are experienced by civilised man and by the primitive savage, it would appear that an illusory presentation must in some way prevent the individual from exercising his "g". In fact, Spearman classifies an

[1] F. Aveling and H. L. Hargreaves, "Suggestibility with and without Prestige in Children", *Brit. Journ. of Psychol.* vol. 12, 1921.

[2] W. Köhler, *Gestalt Psychology*, 1929, p. 125.

[3] R. H. Wheeler, *The Science of Psychology*, 1929, p. 359.

illusion as a type of error due to the violence of some bias which acts as a substitute for insight.[1] In every illusion the quantitative law of retentivity in the form of reproduction or persistence overshadows the laws of eduction. An example of persistence is the Müller-Lyer illusion which can be explained as the merging of one cognised item into another cognised immediately afterwards. For even when the exposure of the figures is extremely brief, introspection distinguishes between two stages in the perceptual operation: first, the entire figures are seen with diffused attention, and then one or both of the stems are seen with concentrated attention, and only in this second period is it that the one stem seems longer than the other, and this apparent greater length of the stem in the "feather head" figure has its origin in the greater lateral extension of the whole "feather head" figure as compared with the "arrow head" figure. This is known as *confluence*,[2] by which is meant that the recurrence of an aspect in one part of a perception facilitates its occurrence in another part. Thus the perception of further extension in the "feather head" parts facilitates our perceiving further extension also in the stem. Spearman indeed does not confine the term "illusion" to sensory perceptions but includes in it all beliefs which are due to an essentially similar mechanism.

QUESTIONS

1. Prepare 8 circular pieces of white paper of the following diameters: $2 - \frac{2}{32}$, $2 - \frac{1}{32}$, 2, $2 + \frac{1}{32}$, $2 + \frac{2}{32}$, $2 + \frac{3}{32}$, $2 + \frac{4}{32}$, $2 + \frac{5}{32}$ inches, also 8 square pieces of white paper, the side of each square being 1·8 inches, also 8 sheets of neutral gray, 9 in. by 7 in., on each of which should be pasted one of the circles and one of the squares, the distance between them being 1 inch. E exposes these sheets one at a time, according to the Constant Method when the steps are unequal (Chapter 24). Each sheet is exposed for about 2 seconds at a constant distance from S, who judges whether the circle is greater,

[1] C. Spearman, *The Nature of "Intelligence"*, 1923, p. 287.

[2] Cf. V. Hazlitt, *Ability*, 1926, p. 13 ff., and A. W. P. Wolters, *The Evidence of our Senses*, 1933, p. 22.

equal or less than the square. Calculate the size of the circle which appears equal to the square.

2. Successive stimulation by two lights at two points A and B not too distant from each other produces an experience of movement from A to B, provided the illumination in each case is of suitable intensity and duration, and the time interval between the two stimulations is also suitable. Do you regard such a phenomenon as an illusion? (This is the well-known *phi phenomenon* of Wertheimer, and the advanced student may consult M. Wertheimer's "Experimentelle Studien über das Sehen von Bewegung", *Zeitsch. f. Psychol.* 61, 1912; also R. H. Wheeler's *The Science of Psychology*, p. 362, 1929, where Korte's Laws of optimum movement are stated.)
3. A white disc, 1 foot in diameter, lies flat on a table with its centre 4 feet from the edge of the table. S, with his eyes 18 inches above the edge of the table, looks at the disc through a slit in a screen and matches the apparent shape of the disc with one of a series of cardboard ellipses graded in eccentricity. It will be found that he will choose one intermediate in shape between the "real shape" which is a circle, and the "perspective shape" which is a particular ellipse. Is this an illusion? (cf. R. H. Thouless, "Phenomenal Regression to the Real Object", *Brit. Journ. of Psychol.* vols. 21 and 22, 1931; "Individual Differences in Phenomenal Regression", *ibid.* vol. 22, 1932; W. Köhler, *Gestalt Psychology*, 1929, p. 76. Thouless calls the apparent shape the "phenomenal shape" and found that Ss differed in their tendency to phenomenal regression to the real object. Women tended to show greater regression than men. Artists showed less regression than the average. Regression was found to correlate positively with age and negatively with intelligence).

References

1. C. E. Seashore, *Measurements of Illusions and Hallucinations in Normal Life*, Studies from the Yale Psychol. Lab. 3, 1895.
2. C. E. Seashore, *The Material-weight Illusion*, Univ. of Iowa Studies in Psychol. 2, 1899.

3. A. Binet, *La Suggestibilité*, 1900.
4. E. B. Titchener, *Experimental Psychology*, vol. 1, Part. 2, 1901.
5. W. H. R. Rivers, "Observations on the Senses of the Todas", *Brit. Journ. of Psychol.* vol. 1, 1905.
6. E. O. Lewis, "The Effect of Practice on the Perception of the Müller-Lyer Illusion", *Brit. Journ. of Psychol.* vol. 2, 1908.
7. C. S. Myers, *Introduction to Experimental Psychology*, 1911.
8. E. O. Lewis, "Confluxion and Contrast Effects in the Müller-Lyer Illusion", *Brit. Journ. of Psychol.* vol. 3, 1919.
9. C. Spearman, "The Origin of Error", *Journ. of General Psychol.* vol. 1, 1928.

CHAPTER 23

STATISTICAL METHODS

§ 1. STATISTICS

The word "statistics" is derived from the Latin *status*, a political state. The word "statist", meaning a politician, is found in *Hamlet*, Act V. Sc. II. (1602). But it was nearly two hundred years later when the word "statistics" appeared. It then signified the exposition of the noteworthy characteristics of a state, but this was mostly in words not in figures. Not until the nineteenth century did the word acquire a narrower meaning, namely, the exposition of the characteristics of a state by numerical methods, *e.g.* vital statistics, poor-law statistics, etc. Finally, about the end of the nineteenth century, the word "statistics" was used not merely with reference to matters of state but quite generally to all quantitative data, whether found in the natural, biological or social sciences. At present, then, those methods which are specially adapted to the elucidation of quantitative data affected by a multiplicity of causes are called statistical methods.[1]

As the psychologist has to deal with highly complicated cases of multiple causation it becomes essential that he should have some acquaintance with statistical terminology. Otherwise, he can neither hope to read psychological literature with pleasure and understanding nor attempt to deal authoritatively with any quantitative data on his own account. Even to the expert statistician discretion and humility are still desirable attributes lest there be a causal circumstance remaining uncontrolled.

Now there are different orders of statisticians, as Woodworth[2] has pointed out. There is on the one hand the mathematician

[1] G. Udny Yule, *An Introduction to the Theory of Statistics*, 1927, pp. 1-5.

[2] R. S. Woodworth in Introduction to H. E. Garrett's *Statistics in Psychology and Education*, 1926.

who invents the method or formula, and on the other the computer who can, when directed, work out the necessary averages or coefficients as demanded by the formulae. In between the two comes the psychologist who "must have a discriminating knowledge of the kit of tools which the mathematician has handed him as well as some skill in their actual use". He may not have an accurate knowledge of how a formula was derived, yet he must know the use and limitations of it.

§ 2. CONTINUOUS SERIES AND DISCONTINUOUS SERIES

If it be necessary to measure the Intelligence Quotient of each child in the elementary schools of a large city it would be found, with rare exceptions, that all the measures would lie somewhere between 60 and 160. Provided the divisions of the scale were sufficiently minute any measure whatever between these limits might be met, *e.g.* 106·7. Thus the series of measurements may be regarded as *continuous*.

It is otherwise when the variable is the size of a family, the size of a school class, salary per week, etc. There can be no school class intermediate in size between 30 and 31 pupils. Such variables are therefore *discontinuous*. Most data which are gathered in psychology belong to continuous series. But there are two ways of regarding the activity of marking when dealing with continuous measures. Usually as in a group test of intelligence, each pupil who gets 8 marks may be regarded as deserving a mark lying somewhere between 8 and 8·99 . . . , for some may have been on the point of completing the next answer when time was called. But sometimes, in such performances as essays or handwriting, 8 marks may be taken at face-value as expressing the judgment of the examiner better than 7 or 9, that is, it is 8 to the nearest whole number.

§ 3. FREQUENCY DISTRIBUTION

After collecting the data, the first task, if the measures are at all numerous, is to classify them in an orderly form. Thus, supposing it is required to find the Intelligence Quotient of all boys aged eleven attending a big elementary school and that the results have to be presented for analysis. The lowest value

might be 71·5 and the highest 139. By choosing a class interval of 5 this range will provide 14 steps. As a rule, less than 10 steps does not provide a detailed enough analysis and more than 20 steps would be cumbersome. In the following table every I.Q. equal to or greater than 70 but less than 75 is entered as belonging to the class interval 70–74·9, and similarly for the other measures.

§ 4. REPRESENTATIVE VALUES

It is impossible to keep in view the individual measurements of an extended series. Instead a measure is sought which represents these measurements. There are three such representative values in use: (1) the average or mean, (2) the median, (3) the mode.

The *Mean* is the measurement obtained by adding up all the

Class Interval	Frequency			
	f	d	fd	fd^2
135–139·9	1	7	7	49
130–134·9	2	6	12	72
125–129·9	3	5	15	75
120–124·9	4	4	16	64
115–119·9	3	3	9	27
110–114·9	5	2	10	20
105–109·9	8	1	8	8
100–104·9	11	0	77	0
95– 99·9	9	- 1	- 9	9
90– 94·9	5	- 2	- 10	20
85– 89·9	5	- 3	- 15	45
80– 84·9	2	- 4	- 8	32
75– 79·9	1	- 5	- 5	25
70– 74·9	1	- 6	- 6	36
	60	..	- 53	482

$$fd = 77 - 53 = 24$$

$$c = \frac{24 \times 5}{60} = 2$$

$$\text{Mean} = 102{\cdot}5 + 2 = 104{\cdot}5.$$

individual measurements and dividing by the number of measures. A little reflection, however, will convince the student of the great saving of time by using the short method recommended for actual practice when the measurements have been arranged in a frequency distribution. It is customary to assume a mean at the middle point of the class interval of greatest frequency. Thus in the present example the assumed mean is 102·5. The *d* column gives the deviations of the middle points of the class intervals from this assumed mean in units of class interval. To get the exact mean, multiply each *d* by its *f* and find the algebraic sum which is 24. On dividing this sum by the number of cases and multiplying by the size of the class interval, the correction *c*, which should be added to the assumed mean in order to get the exact mean, is obtained. Thus the exact mean is 104·5.

The *Median* is the point above which and below which are 50 per cent of the measures. In the above example it may be found by counting off one half of the measures, namely, 30. Twenty-three measures bring the score to 100, thus the median is

$$100 + \frac{7}{11} \times 5 = 103{\cdot}2.$$

The *Mode* is that measurement which occurs most often in the series. It is rarely of use in psychology and will here be neglected.

COMPARISON OF MEAN AND MEDIAN.—The mean takes into account the size of every measure. This is an advantage if the extreme measures are valid, for then they should be permitted to contribute their weight. But this is a disadvantage if the extreme measures are suspicious, for they may greatly affect the value of the mean.

The mean in general is more reliable than the median, as the standard deviation (cf. § 6) of the latter is 25 per cent more if the distributions are approximately normal. The mean is also to be preferred when the data have to be subjected to further analysis, *e.g.* the computation of the product moment coefficient of correlation. Further, in combining the representative values

of several series the mean has an advantage for it can be obtained algebraically from the separate means, while to obtain the median it is necessary to combine all of the series into one and calculate afresh.

§ 5. MEASURES OF VARIABILITY

In addition to the representative value, there is need of another fundamental concept, namely, the variability of the scores. There are four ways of measuring the variability or dispersion.

(1) *The Range.*—As an illustration consider the two distributions:

5	15	25	35	45
23	24	25	26	27

Both have the same number of cases, namely 5, and the same mean, namely 25. But the range or spread in the first series is 40 and in the second only 4. The range then is a rough measure depending only on the two extreme scores, but it may be useful when only a few measurements are available. In the example of § 3 the range of Intelligence Quotients is $139 - 71{\cdot}5 = 67{\cdot}5$.

(2) *The Average Deviation* (A.D.).—For a simple ungrouped series the average deviation is obtained by adding the absolute values of the deviations from the mean, disregarding sign, and dividing by their number. For the example shown it is 12.

X	D	D^2
45	20	400
35	10	100
25	0	0
15	10	100
5	20	400
	60	1000

Mean = 25.

$$\text{Average Deviation} = \frac{\Sigma D}{n} = \frac{60}{5} = 12.$$

$$\text{Standard Deviation} = \sqrt{\frac{\Sigma D^2}{n}} = \sqrt{\frac{1000}{5}} = \sqrt{200} = 14 \text{ (approx.)}.$$

It must be distinctly understood that this example, with n as ridiculously low as 5, is merely taken for illustration. No one would attach any importance to a measure of variability obtained from such a small number of scores.

For small samples, when n is less than 30, the estimated value of the average deviation or the standard deviation may differ considerably from the true value, and other methods of estimating the significance of statistics for such cases will be discussed later.

For a grouped distribution as in § 4, let A_m be the middle point of the interval in which the mean M lies, let N_a and N_b denote the number of measurements above and below the mean, and let h be the size of the class interval. Then [1]

$$\text{A.D.} = \frac{\Sigma|fd|h + (A_m - M)(N_a - N_b)}{n}$$

$$= \frac{130 \times 5 + (102{\cdot}5 - 104{\cdot}5)(26 - 34)}{60} = 11{\cdot}1.$$

(3) *Standard Deviation* (σ).—For a simple ungrouped series the deviations from the mean are squared, added together and averaged, and the square root of this average is the standard deviation of the sample, or σ_{dis}.

For the example given it is 14.

For a grouped series, as in § 4,

$$\sigma_{\text{dis.}} = \left(\sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2}\right)h$$

$$= \left(\sqrt{\frac{482}{60} - \left(\frac{24}{60}\right)^2}\right)5 = 14.$$

The standard deviation is a more reliable measure of variability than the A.D. It is less affected by fluctuations in sampling, and as in the case of the mean its algebraic properties makes it useful when the results from several series have to be combined.

(4) *The Quartile Deviation.*—If Q_1, the lower quartile, be determined so that one-quarter of all the measures are less than Q_1 and three-quarters greater. And if Q_3, the upper quartile, be

[1] $\Sigma|fd|$ is the *arithmetic* sum of the fd's. In this example all measurements within the interval where M lies are assumed to be below M.

determined so that one-quarter of all the measures are greater and the other three-quarters less. Then the range Q_1 to Q_3 is the *inter-quartile range* and includes the middle half of the measures. The *semi-inter-quartile range* is the quartile deviation Q so that

$$Q = \frac{Q_3 - Q_1}{2}.$$

In § 4, Q_1, the 25th percentile, is $95 + \frac{1}{9} \times 5 = 95{\cdot}5$ (since $\frac{N}{4} = 15$).

Q_2, the median, or 50th percentile, has already been calculated.

Q_3, the 75th percentile, is $110 + \frac{3}{5} \times 5 = 113$ (since $\frac{3N}{4} = 45$).

(5) Pearson's Coefficient of Variation (V).—The measures of variability thus far considered are expressed in the units of the variable and are independent of the magnitude of the mean. But sometimes we wish to compare the variability of a group in one activity with its variability in another where the scale units and the mean are quite different.

This can be effected by Pearson's coefficient of variation or

$$V = \frac{100\sigma}{M}.$$

The mean cost of a suit of clothes for adult males in a given community might be £4 and the standard deviation £2, whereas the mean cost of the houses in which they live might be £800 and the standard deviation £300. Although the absolute variability in cost of house is 150 times greater than that in cost of a suit, yet the relative variabilities for house and suit respectively are $37\frac{1}{2}$ and 50. So that the members of this community are actually less variable in their choice of house than in their choice of a suit.

Example.—Compare the variabilities of a class in examinations in English and in Arithmetic. For English, $M_1 = 80$ marks; $\sigma_1 = 10$ marks. For Arithmetic, $M_2 = 40$ marks; $\sigma_2 = 10$ marks.

§ 6. MEASURES OF RELIABILITY

In the illustration of § 4 the group of sixty is only one sample of boys of eleven taken from a particular school. The question

arises: How truly does a representative value, *e.g.* the mean of that sample, 104·5, apply with reference to all boys of eleven throughout the country? And further: How truly does a measure of variability, *e.g.* the standard deviation for that sample, 14, represent that obtained when all boys of eleven are included? Before answering it is necessary to consider errors of two kinds:

(1) *Variable Errors.*—The sample taken may have had long practice with similar tests, or may have been specially coached, or it may be fatigued, or inattentive, or the tester may not have administered the tests according to the printed directions. Unless such factors can be controlled it would be idle to expect the values obtained to have universal significance.

(2) *Sampling Errors.*—The sample may contain a greater percentage of bright scholars than would be found in a group representative of the whole country, that is, the sample is not truly representative. In fact, the value obtained for the mean 104·5, suggests this. In such a case no formula would enable us to calculate the true mean, or true standard deviation. But if the sample can be regarded as a random sample, then Yule has proved, and it can also be verified in practice, that the reliability of a mean increases in proportion to the square root of the number of measurements. It is also clear that the mean is likely to be more reliable the more closely the individual measurements are clustered round the mean, that is, the smaller $\sigma_{\text{dis.}}$ is.

Thus the reliability of a mean can be expressed in terms of the standard deviation of the mean $\sigma_{\text{av.}}$

$$\therefore \quad \sigma_{\text{av.}} = \frac{\sigma_{\text{dis.}}}{\sqrt{\text{N}}}.$$

In applying this formula it is not necessary to assume that the distribution curve is normal. Even when it is not normal it will still be found that for a large number of samples the deviations of the mean of each sample from the true mean will tend to follow the normal curve provided the size of each sample is large. In the laboratory, however, small samples are often

considered and special methods have been devised for determining the reliabilities (cf. § 7).

There are other formulae which can only be used when the distribution curve is normal or approximately normal. Thus the reliability of the median can be expressed in terms of its standard deviation $\sigma_{\text{Med.}}$:

$$\sigma_{\text{Med.}} = 1{\cdot}2533 \frac{\sigma_{\text{dis.}}}{\sqrt{\text{N}}}.$$

Likewise the reliability of an obtained $\sigma_{\text{dis.}}$ is given by σ_σ and

$$\sigma_\sigma = 0{\cdot}7071 \frac{\sigma_{\text{dis.}}}{\sqrt{\text{N}}}.$$

The Reliability of the Difference between Two Means.—In § 4 the mean I.Q. for boys of eleven was 104·5. If the corresponding value for girls of eleven from the same school were 106·5 or 102·5, one could hardly conclude without further analysis that there was a significant difference in I.Q. between boys and girls of eleven. The samples taken, being from one particular school, are not sufficiently representative. It would be necessary to take a large random sample of boys and girls of eleven and apply the formula

$$\sigma_{\text{diff.}} = \sqrt{\sigma^2_{\text{av.}_1} + \sigma^2_{\text{av.}_2}},$$

where $\sigma_{\text{diff.}}$ is the standard deviation of the difference between the two means. Unless the observed difference D between the means is at least three times as great as $\sigma_{\text{diff.}}$ it is not usual to rely on the difference being significant. Thus supposing the means for 300 boys and 400 girls were 104·5 and 105·5 respectively, and the corresponding standard deviations of the two distributions 10 and 15 respectively then

$$\sigma_{\text{diff.}} = \sqrt{\frac{100}{300} + \frac{225}{400}} = 0{\cdot}95$$

and

$$\frac{\text{D}}{\sigma_{\text{diff.}}} = \frac{105{\cdot}5 - 104{\cdot}5}{0{\cdot}95} = 1.$$

From the following table it can be seen that the chances of the girls being superior to the boys would be only 84 in 100 and hence not very significant. The first row gives $\frac{D}{\sigma_{\text{diff.}}}$ and the second the chances in 100.

0	0·2	0·4	0·6	0·8	1·0	1·2	1·4	1·6	1·8	2·0	2·2	2·4	2·6	2·8	3·0
50	58	65	73	79	84	88	92	94	96	98	98·6	99·2	99·5	99·8	99·9

In the last example the two series of measurements, namely, those for boys and for girls, were assumed to be independent of each other. But sometimes the two series of measurements are correlated, then the appropriate formula must be employed:

$$\sigma_{\text{diff.}} = \sqrt{\sigma^2_{\text{av.}_1} + \sigma^2_{\text{av.}_2} - 2r\sigma_{\text{av.}_1} \cdot \sigma_{\text{av.}_2}},$$

where r is the coefficient of correlation between the two series.

§ 7. STATISTICS FOR SMALL SAMPLES

It was pointed out by *Student* in 1908 that the standard deviation, σ, of a population cannot accurately be estimated from the small sample which is available, and a method of finding the significance of the mean in such a case was developed.[1] In 1925 R. A. Fisher's *Statistical Methods for Research Workers* appeared, and Fisher's methods of dealing with small samples are now becoming widely known in psychological laboratories. A brief account of Fisher's methods of finding the significance (1) of the mean and (2) of the difference between means will now be given.

SIGNIFICANCE OF THE MEAN.—If $x_1, x_2, \ldots x_n{}^1$ is a small sample of n^1 values and if s is the *estimated* standard deviation, and $\bar{x}$ is the mean, then the *estimated* standard deviation of the mean is $\frac{s}{\sqrt{n^1}}$. Hence the chances of the mean being zero will depend on the value of

$$t = \frac{\bar{x}}{\frac{s}{\sqrt{n^1}}}.$$

[1] *Student*, "On the Probable Error of a Mean," *Biometrika*, vol. 6, 1908.

To find s Fisher uses the formula

$$s = \sqrt{\frac{S(x-\bar{x})^2}{n^1 - 1}} \quad \text{instead of} \quad \sigma = \sqrt{\frac{S(x-\bar{x})^2}{n^1}}.$$

Consequently $$\frac{s^2}{n^1} = \frac{1}{n^1(n^1-1)} S(x-\bar{x})^2.$$

For small samples this increases the value of the deviation.

A table of t values is given by Fisher which enables us to determine the significance of the mean. In the table $n = n^1 - 1$. To be significant, the value of P, the probability of falling outside the range $\pm t$, should at least be less than 0·05.

Example.—Six professional golfers who returned the best scores at a large tournament were requested to test the standard size ball against one slightly larger. After some preliminary practice with the larger ball it was decided that each should play a morning and an afternoon round. Each started with a standard ball in the morning and with the larger ball in the afternoon. Each changed from one type of ball to the other after playing each hole. From the figures appended find the probability of one type of ball being superior.

Player	Standard Ball	Larger Ball	x Difference	$x-\bar{x}$	$(x-\bar{x})^2$	$x-\bar{x}$	$x^1-\bar{x}^1$	$(x-\bar{x})^2$	$(x^1-\bar{x}^1)^2$
A	70	73	+3	+1	1	−3	−2	9	4
B	71	72	+1	−1	1	−2	−3	4	9
C	72	72	0	−2	4	−1	−3	1	9
D	74	76	+2	0	0	+1	+1	1	1
E	75	73	−2	−4	16	+2	−2	4	4
F	76	84	+8	−6	36	+3	+9	9	81
Mean	73	75	+2		58 ÷ 6			28	108

$$\bar{x} = \frac{1}{n^1} S(x) = 2; \qquad \frac{s^2}{n^1} = \frac{1}{n^1(n^1-1)} S(x-\bar{x})^2 = \frac{1}{30} \times 58;$$

$$t = \bar{x} \div \sqrt{\frac{s^2}{n^1}} = 1{\cdot}44; \qquad n = n^1 - 1 = 5.$$

Hence from the t table P is about 0·2 and the result is not very significant.

SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS.—If $x_1, x_2, \ldots x_{n_1+1}$ and $x^1{}_1, x^1{}_2, \ldots x^1{}_{n_2+1}$ be two samples, the significance of the difference between their means $\bar{x}$ and $\bar{x}^1$ depends on the value of

$$t = \frac{\bar{x} - \bar{x}^1}{\sigma_{\text{diff.}}}.$$

Now if the true σ of the population were known, we have already seen that $\sigma^2_{\text{diff.}}$ would be

$$\frac{\sigma^2}{n_1 + 1} + \frac{\sigma^2}{n_2 + 1}.$$

But as it is not known Fisher forms an estimate s of it by pooling the sums of squares from the two samples and dividing by the total number of the degrees of freedom contributed by them, namely $n_1 + n_2$. The necessary statistics then must be

$$\bar{x} = \frac{1}{n_1 + 1}\text{S}(x) \qquad \bar{x}^1 = \frac{1}{n_2 + 1}\text{S}(x^1);$$

$$s^2 = \frac{1}{n_1 + n_2}\left\{\text{S}(x - \bar{x})^2 + \text{S}(x^1 - \bar{x}^1)^2\right\};$$

$$t = \frac{\bar{x} - \bar{x}^1}{s}\sqrt{\frac{(n_1 + 1)(n_2 + 1)}{n_1 + n_2 + 2}};$$

$$n = n_1 + n_2.$$

Example.—Suppose the figures in the first two columns of the last table had been obtained using different players for the two balls, what is the probability of one type of ball being superior?

On calculating the necessary statistics, using the last two columns in the table, it is seen that

$$s^2 = \frac{1}{10}(28 + 108) = 13{\cdot}6,$$

and since $\bar{x} - \bar{x}^1 = 2$

$$\therefore \quad t = 0{\cdot}94.$$

For $n=10$, the resulting value of P from the t table is between 0·4 and 0·3, The difference then is clearly not significant, and as would be expected, the result is far more significant when the same players use both balls.

ABBREVIATED TABLE OF t

n	P=0·9	0·8	0·7	0·6	0·5	0·4	0·3	0·2	0·1	0·05	0·02	0·01
1	0·16	0·32	0·51	0·73	1·00	1·38	1·96	3·08	6·31	12·71	31·82	63·66
5	0·13	0·27	0·41	0·56	0·73	0·92	1·16	1·48	2·01	2·57	3·36	4·03
10	0·13	0·26	0·40	0·54	0·70	0·88	1·09	1·37	1·81	2·23	2·76	3·17
20	0·13	0·26	0·39	0·53	0·69	0·86	1·06	1·32	1·72	2·09	2·53	2·84
30	0·13	0·26	0·39	0·53	0·68	0·85	1·05	1·31	1·70	2·04	2·46	2·75
∞	0·13	0·25	0·39	0·52	0·67	0·84	1·04	1·28	1·64	1·96	2·33	2·58

§ 8. REGRESSION AND CORRELATION

In addition to computing a representative value such as the mean variate, or a measure of the variability of the scores, or a measure of the reliability of the mean; and in addition to comparing two variates as regards their means or variabilities, or determining the reliability of the difference between their means, it is often essential in the social sciences to ascertain how, and to what extent, the two variates are related. For this purpose it is necessary to become familiar with two important and distinct concepts, namely, the *regression coefficient* and the *correlation coefficient*. Fisher [1] points out: "The idea of regression is usually introduced in connexion with the theory of correlation, but it is in reality a more general, and, in some respects, a simpler idea; and the regression coefficients are of interest and scientific importance in many classes of data where the coefficient of correlation, if used at all, is an artificial concept of no real utility".

Sir Francis Galton [2] introduced the terms regression and co-relation and devised methods of measuring them in his studies of inheritance. Truly a fascinating chapter in the history of

[1] R. A. Fisher, *Statistical Methods for Research Workers*, 1932, p. 120.

[2] F. Galton, "Regression towards Mediocrity in Hereditary Stature", *Journ. of Anthrop. Inst.* vol. 15, 1886, and "Family Likeness in Stature", *Proc. Roy. Soc.* vol. 40, 1886.

scientific progress! By regression he meant the tendency of a son's predicted height to be nearer the mean of males than the height of his father. Thus the son of a tall father would be expected to be shorter, and the son of a short father to be taller. In each case the measure of regression gives the expected height.

By co-relation was meant the relation between the heights of mid-parents and heights of sons, where mid-parent height is the mean height of father and mother. The terms regression and correlation continue to be used, quite apart from problems of inheritance, in statistical work generally, although as Yule[1] points out, the idea of a regression towards a more or less stationary mean is quite inapplicable when the variables are different in kind.

Karl Pearson continued Galton's work and, in a monumental series of researches developed exact methods for regressional and correlational studies.

Charles Spearman, in addition to developing new methods, has by their use gone furthest towards setting quantitative psychology on firm foundations.

By the aid of such methods the computing of the correlation coefficient becomes a simple matter, but it is not difficult to find cases of psychologists applying the operation to unsuitable data, apparently without realising its limitations or ascertaining from a study of the regressions if the operation was justifiable. In recent years the classic researches of R. A. Fisher are proving a potent aid to psychologists in many ways, especially his study of regression and his method of variance. For example, a vocational counsellor may ascertain what reliance should be placed on the predicted score for the dependent variate, such as musical aptitude, as deduced from the actual test score in terms of the independent variate.

Example of Regression.—A simple illustration of regression is Galton's famous study of the heights of fathers and sons. To simplify the treatment we may assume for our sample that (1) each father has only one son, so as to avoid tabulating the father's height more than once, (2) the average stature is 68 in. for fathers and also for sons, a value which is sufficiently exact

[1] G. Udny Yule, *Theory of Statistics*, 1922, p. 177.

for our purpose, and (3) the statures vary between 60 and 76 in., neglecting the few cases outside these limits so as to get a more compact diagram in illustration. Now consider all fathers whose heights are between 60 and 61 inches and find the mean height of their sons and mark it with a cross in the diagram. Similarly for the sons of fathers whose heights are between 61 and 62 in. and so on. In this way a series of crosses is obtained. Next

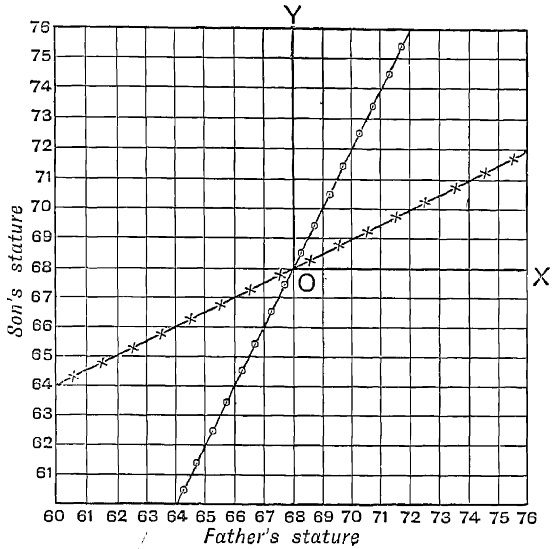

FIG. 31.

consider all sons whose heights are between 60 and 61 in. and find the mean height of their fathers and mark it with a circle, and similarly the mean height of fathers whose sons are between 61 and 62 in. and so on. In the figure the series of crosses and circles lie on two straight lines, but in an actual sample of fathers and sons the crosses and lines would only show a tendency to lie on straight lines. In that case the regression lines are the "lines of best fit". When the crosses or circles lie close to such a line it can easily be drawn but as they usually show a certain scatter there is need of some method to determine the equations of the best-fitting lines. In the above example, then,

the regression is said to be *linear*. When the circles or crosses do not approximately fit such a line the regression is not linear. Such cases are beyond the scope of this chapter.

It is an easy exercise to consider what would have been the significance (1) were the two regression lines coincident, or (2) if one were coincident with the OX-axis and the other with the OY-axis.

The equation of the first regression line through the crosses with reference to the axes through the means may be written

$$y = b_x x.$$

where b_x is the regression coefficient of y on x. This equation is only to be used in order to predict the probable deviation from the mean of the height of a son given that his father's height is x.

Similarly the equation of the second regression line

$$x = b_y y.$$

should only be used in order to predict the probable deviation from the mean of the height of a father given that his son's height is y and the regression coefficient of x on y is b_y.

It is easy to express the above equations to the regression lines in *score form*, so that the predicted heights can be obtained directly. They are

$$\mathrm{Y} = a + b_x(x - \bar{x}), \quad \text{where } a = \bar{y};$$
$$\mathrm{X} = a^1 + b_y(y - \bar{y}), \quad \text{where } a^1 = \bar{x}.$$

Thus in the above example the equations of the regression lines are

$$\mathrm{Y} = 68 + \frac{1}{2}(x - 68),$$

$$\mathrm{X} = 68 + \frac{1}{2}(y - 68),$$

reducing to

$$\mathrm{Y} = 34 + \frac{x}{2},$$

$$\mathrm{X} = 34 + \frac{y}{2}.$$

But it is important to realise that, in this example, we have made use of Galton's data and therefore *know* that $\bar{y}=\bar{x}=68$ in. approximately, and that $b_x=b_y=0{\cdot}5$ approximately. In actual research, however, the equation of the regression line

$$\mathrm{Y}=a+b(x-\bar{x})$$

has to be obtained from the data furnished by the sample, and thus the estimated values of a and b are subject to the errors of random sampling. According to Fisher's notation, if the true regression formula obtained from an infinity of observations is

$$\mathrm{Y}=\alpha+\beta(x-\bar{x}),$$

then the significance of the difference between a and any hypothetical value α can be obtained from the t table by calculating

$$t=\frac{(a-\alpha)\sqrt{n^1}}{s}, \quad n=n^1-2,$$

where n^1 is the number of values in the sample. Similarly the significance of the difference between b and any hypothetical value β is obtained by calculating

$$t=\frac{(b-\beta)\sqrt{\mathrm{S}(x-\bar{x})^2}}{s}, \quad n=n^1-2.$$

For the best estimate of the value of σ^2, where σ is the standard error of y for a given value of x, is given by

$$s^2=\frac{1}{n^1-2}\mathrm{S}(y-\mathrm{Y})^2$$

and is found by summing the squares of the deviations of y from its calculated value Y, and dividing by n^1-2. (Proofs of these formulae are given in R. A. Fisher's *Statistical Methods for Research Workers*, 1932, pp. 123-127.)

§ 9. THE MEASUREMENT OF CORRELATION

PEARSON'S PRODUCT-MOMENT METHOD.—This method uses the actual marks or scores. Three cautions are necessary: (1) It

would be futile to calculate a coefficient of correlation with only seven subjects, it is only done here as an illustration. (2) In actual practice Pearson's product-moment coefficient is calculated by shorter methods such as the one shown later in this section when measuring the relation between the sense of pitch and tonal memory. (3) It is not to be assumed that the coefficient here calculated represents the true value to be expected.

Pupil	Arithmetic	Algebra	x	y	xy +	xy −	x^2	y^2
A	16	17	+6	+6	36		36	36
B	14	10	+4	−1		4	16	1
C	11	15	+1	+4	4		1	16
D	9	12	−1	+1		1	1	1
E	8	13	−2	+2		4	4	4
F	6	7	−4	−4	16		16	16
G	5	6	−5	−5	25		25	25
	69	80	..	..	+72		99	99

The table gives the scores of seven pupils in arithmetic and in algebra. It will be seen that the average scores in the two tests are about 9·9 and 11·4, but in calculating the deviations x and y from the averages it is quicker to regard the nearest whole numbers, namely, 10 and 11, as the averages. The required coefficient of correlation is

$$r=\frac{S(xy)}{\sqrt{S(x^2)\cdot S(y^2)}}=\frac{72}{99}=0{\cdot}73.$$

The probable error of r is given by

$$PE_r=0{\cdot}6745\frac{1-r^2}{\sqrt{n}}.$$

The error introduced by such approximation is small and, if necessary, the exact value of r may be found from the formula

$$r=\frac{S(xy)-ne_1e_2}{\sqrt{[S(x^2)-ne_1{}^2][S(y^2)-ne^2{}_2]}}$$

$$=\frac{72-7(+0{\cdot}1)\cdot(-0{\cdot}4)}{\sqrt{[99-7(0{\cdot}1)^2][99-7(0{\cdot}4)^2]}}=0{\cdot}73 \text{ as before.}$$

S(xy) denotes the algebraic sum of the products of the deviations x and y. S(x^2) and S(y^2) denote the sums of the squares of the x and y deviations respectively. e_1 is the error in taking 10 instead of 9·9, and e_2 the error in taking 11 instead of 11·4.

Spearman's Method of Ranks.—In this method[1] it is only necessary to know the rank order in each test. D denotes the difference between each pupil's ranks in the two tests, and ρ denotes the coefficient of correlation by this method.

Pupil	Arithmetic	Algebra	D	D^2
A	1	1	0	0
B	2	5	3	9
C	3	2	1	1
D	4	4	0	0
E	5	3	2	4
F	6	6	0	0
G	7	7	0	0
				14

$$\rho = 1 - \frac{6\,S(D^2)}{n(n^2-1)}$$
$$= 1 - \frac{6 \times 14}{7 \times 48}$$
$$= 0{\cdot}75$$

The probable error of ρ is given by

$$PE_\rho = \frac{0{\cdot}7063(1-\rho^2)}{\sqrt{n}}.$$

In practice the values obtained by this method agree fairly closely with those given by the product-moment method. Tables are given in books on statistics for converting ρ into r, but the correction is so small in comparison with other possible errors that it may well be negelcted here. It is, of course, possible to obtain a correlation of unity by this method, although it would not exactly be unity by the product-moment method, *e.g.* the best pupil might have a lead in arithmetic of twenty marks but in algebra of only one. Spearman, however, compared the two methods and shows that the method of

[1] C. Spearman, "The Proof and Measurement of Association between Two Things", *Amer. Journ. of Psychol.*, vol. 15, 1904.

ranks may possess certain advantages, the chief being the reduction of the "accidental error". For errors in scoring the outlying exceptional cases may have a serious effect when the product-moment method is used.

SPEARMAN'S FOOT-RULE METHOD.—This is a quick method of finding an approximate value of the coefficient of correlation between two rank orders.

Pupil	Arithmetic	Algebra	G
A	1	1	..
B	2	5	..
C	3	2	1
D	4	4	..
E	5	3	2
F	6	6	..
G	7	7	..
			3

If G is the gain in rank, the coefficient of correlation, R, furnished by this method is

$$R = 1 - \frac{6S(G)}{n^2 - 1}$$

$$= 1 - \frac{6 \times 3}{48} = 0{\cdot}625.$$

A rough estimate of its probable error is

$$PE_R = \frac{\cdot 43}{\sqrt{n}}.$$

The following abridged table, based on the formula

$$r = 2 \cos \frac{\pi}{3}(1 - R) - 1,$$

is for the purpose of converting R values into r values.

TABLE FOR CONVERTING R INTO r

R	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50
r	0·09	0·18	0·26	0·34	0·41	0·49	0·55	0·62	0·68	0·73
R	0·55	0·60	0·65	0·70	0·75	0·80	0·85	0·90	0·95	1·00
r	0·78	0·83	0·87	0·90	0·93	0·96	0·97	0·99	1·00	1·00

METHOD OF COMPUTING CORRELATION FROM A FREQUENCY TABLE.—To simplify the arithmetic a small group of only 30 girls will be selected for illustration. They were given Seashore's Sense of Pitch and Tonal Memory Tests and the marks were distributed as in the table. Thus one girl had a mark between 90 and 100 in Pitch and between 40 and 50 in Tonal Memory.

Let f_x be the frequency of a Pitch column,

f_y be the frequency of a Tonal Memory row,

f_{xy} be the frequency of a cell common to such a column and row,

d_x and d_y the deviations in class intervals from the assumed means for the two tests.

It can be shown that the product-moment formula may be transformed to one containing the frequency notation, namely,

$$r = \frac{\Sigma f_{xy} d_x d_y - \dfrac{(\Sigma f_x d_x)(\Sigma f_y d_y)}{n}}{\sqrt{\left[\Sigma f_x d_x^2 - \dfrac{(\Sigma f_x d_x)^2}{n}\right]\left[\Sigma f_y d_y^2 - \dfrac{(\Sigma f_y d_y)^2}{n}\right]}}.$$

In the present example it is easily seen that

$$\Sigma f_x d_x = -10; \quad \Sigma f_y d_y = -19; \quad \Sigma f_x d_x^2 = 200; \quad \Sigma f_y d_y^2 = 221.$$

The quantity $\Sigma f_{xy} d_x d_y$ is the result of multiplying each cell frequency by its d_x and d_y and then adding all the products formed. But it is quicker to multiply the cell frequencies in a particular column by the appropriate d_y values, add the results

and multiply this by the d_x value of the column. Thus in the column 70–80 the quantities to be added together are 5, 4 and 0. Their sum is multiplied by the appropriate d_x value, which is 3. This column, therefore, contributes 27 towards the total $\Sigma f_{xy}d_xd_y$.

SENSE OF PITCH

Tonal Memory	10	20	30	40	50	60	70	80	90	100		fy	dy	$fy\,dy$	$fy\,dy^2$
90–100					(5) 1			(5) 1				2	5	10	50
80–90								(4) 1				1	4	4	16
70–80			(3) 1		(3) 1							2	3	6	18
60–70									(2) 1			1	2	2	4
50–60						(1) 1	(1) 1					2	1	2	2
40–50		(0) 1	(0) 1		(0) 1			(0) 1		(0) 1		5	0	0	0
30–40	(−1) 1		(−1) 1				(−1) 1		(−1) 1			4	−1	−4	4
20–30	(−2) 1		(−4) 2	(−2) 1	(−2) 1							5	−2	−10	20
10–20		(−3) 1		(−3) 1	(−3) 1							3	−3	−9	27
1–10	(−4) 1	(−8) 2		(−4) 1	(−4) 1							5	−4	−20	80
														−19	221
fx	3	4	5	3	6	1	2	3	2	1		30		$\Sigma fydy$	$\Sigma fydy^2$
dx	−4	−3	−2	−1	0	1	2	3	4	5					
$fxdx$	−12	−12	−10	−3	0	1	4	9	8	5	−10	$\Sigma fxdx$			
$fxdx^2$	48	36	20	3	0	1	8	27	32	25	200	$\Sigma fxdx^2$			
	32	33	4	9	0	1	0	27	4	0	110	$\Sigma fxydxdy$			

FIG. 32.

$$r = \frac{110 - \dfrac{-10 \times -19}{30}}{\sqrt{\left(200 - \dfrac{100}{30}\right)\left(221 - \dfrac{361}{30}\right)}} = 0{\cdot}51.$$

§ 10. THE SIGNIFICANCE OF CORRELATION WITH SMALL SAMPLES

With small samples of n^1 values the observed value of r may be very different from the true value, and consequently this would also be the case for the standard error $\dfrac{1-r^2}{\sqrt{n^1}}$ (or $\dfrac{1-r^2}{\sqrt{n^1-1}}$ as it is sometimes expressed, cf. § 7). The significance of the result may therefore be deceptive. As most correlational researches in psychology involve small samples of about 50

values or less, it is essential for the research worker to become acquainted with the more accurate methods of estimating significance developed by R. A. Fisher. Instead of applying the misleading standard error $\sigma_r = \dfrac{1 - r^2}{\sqrt{n^1 - 1}}$ which would yield

$$t = \frac{r}{\sigma_r} = \frac{r}{1 - r^2} \cdot \sqrt{n^1 - 1},$$

Fisher[1] takes the standard error as $\dfrac{\sqrt{1 - r^2}}{\sqrt{n^1 - 2}}$.

Thus

$$t = \frac{r}{\dfrac{\sqrt{1 - r^2}}{\sqrt{n^1 - 2}}} = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{n^1 - 2}, \quad n = n^1 - 2.$$

On looking up the values of t and n in the t table the significance P may be determined.

§ 11. PARTIAL CORRELATION

In the above example it might be expected that the factor "g" would tend to produce correlation between the two tests. A problem of partial correlation would be to find the correlation between the two variables, sense of pitch and tonal memory, when the influence of the other variable "g" had been eliminated or when the influence of "g" had been held constant. One method would be to use subjects possessing an equal measure of "g" as determined by a reliable test. But the usual method is to give the same subjects a test of "g" and then employ the formula:

$$r_{ab \cdot c} = \frac{r_{ab} - r_{ac} \cdot r_{bc}}{\sqrt{1 - r^2_{ac}}\sqrt{1 - r^2_{bc}}},$$

which gives the correlation between two variables a and b when the third variable c is held constant.

In the above example the correlation between the sense of

[1] R. A. Fisher, *op. cit.* p. 171.

pitch test and an intelligence test was found to be 0·4, and that between the tonal memory test and the intelligence test 0·3. As the observed correlation between the pitch and memory tests was 0·51, it follows that the partial correlation between sense of pitch and tonal memory when general intelligence is held constant is given by

$$r_{pm \cdot g} = \frac{r_{pm} - r_{pg} \cdot r_{mg}}{\sqrt{1 - r^2_{pg}}\sqrt{1 - r^2_{ma}}} = \frac{0{\cdot}51 - (0{\cdot}4 \times 0{\cdot}3)}{\sqrt{1 - (0{\cdot}4)^2}\sqrt{1 - (0{\cdot}3)^2}} = 0{\cdot}44.$$

§ 12. SPEARMAN'S CORRECTION OF A COEFFICIENT OF CORRELATION FOR ATTENUATION

Let an attempt be made to compose two equivalent analogies tests a_1 and a_2, and two equivalent absurdities tests b_1 and b_2, and let a group of 50 eleven-year-old boys be given the tests a_1 and b_1 on one day and a_2 and b_2 on the next. Now if the reliability coefficients $r_{a_1a_2}$ and $r_{b_1b_2}$ are as high as 0·7 it may be concluded that the errors of measurement are not large enough to preclude the tests from further consideration. It is now required to find the correlation between the analogies and absurdities tests.

$r_{a_1b_1}$ or $r_{a_2b_2}$ cannot be regarded as the true measure of this correlation, for each will be subject to "errors of observation" due to slight changes of procedure on the part of the experimenter and fluctuations in the behaviour of the subjects; and although these variations may not affect the average score in each test as they will tend to cancel each other, yet they will increase the standard deviations of the distributions and consequently decrease the coefficient of correlation.

If it be assumed that the errors in any two of the above tests are uncorrelated, and also that the errors in any test are uncorrelated with the scores in any test, then the best available estimate of the correlation between the analogies and absurdities tests is given by the formula

$$r_{ab} = \frac{(r_{a_1b_1} \cdot r_{a_1b_2} \cdot r_{a_2b_1} \cdot r_{a_2b_2})^{\frac{1}{4}}}{(r_{a_1a_2} \cdot r_{b_1b_2})^{\frac{1}{2}}}.$$

Another formula which dispenses with the assumption that the errors in the two tests given on the same day are uncorrelated is

$$r_{ab} = \frac{\sqrt{r_{a_1b_2} \cdot r_{a_2b_1}}}{\sqrt{r_{a_1a_2} \cdot r_{b_1b_2}}}.$$

Even if only one form for each of the two tests is available, an approximate correction for attenuation may be obtained from the following formula, provided the reliability coefficient of each test is known:

$$r_{ab} = \frac{r_{a_1b_1}}{\sqrt{r_{a_1a_2} \cdot r_{b_1b_2}}}.$$

§ 13. EFFECT OF LENGTHENING A TEST ON ITS RELIABILITY

It is sometimes necessary to increase the reliability coefficient of a test by lengthening it. Thus if the application of two equivalent forms of an Analogies Test yielded a reliability coefficient of 0·75, what would be its reliability if it were doubled or trebled in length?

If we assume that the items added are similar in nature to the original the required reliabilities may be calculated by means of the Spearman-Brown Prophecy Formula,

$$\mathrm{R} = \frac{nr}{1+(n-1)r},$$

where r is the original reliability coefficient.

Thus substituting $n=2$, $\mathrm{R}=\cdot857$
and substituting $n=3$, $\mathrm{R}=\cdot90$

In practice, of course, there are several obvious reasons why a test cannot be lengthened indefinitely.

The standard error of the above formula may be written[1]

$$\sigma_n = \frac{n(1+r)}{r\sqrt{\mathrm{N}}}$$

assuming r to be obtained from the scores of N individuals.

[1] E. A. Cureton, "The Standard Error of the Spearman-Brown Formula . . .", *Journ. of Educ. Research*, vol. 24, April 1933.

§ 14. INFLUENCE OF SELECTION OF CORRELATION

LET r_{12} be the correlation between two variables x_1 and x_2 as determined from a large unselected group, and let σ_1 be the variability of x_1. Then if selected values of x_1 are chosen so that the variability of x_1 is reduced to s_1, the resulting correlation R_{12} will be given by the formula[1]

$$R_{12} = \frac{s_1}{\sigma_1} \frac{r_{12}}{\sqrt{1 - r_{12}^2 + r_{12}^2\left(\frac{s_1}{\sigma_1}\right)^2}}.$$

Thus suppose the correlation between an intelligence test and a test in arithmetic for a large unselected group of boys aged eleven years be 0·6, and that all boys whose intelligence quotients are greater than 120 are selected from this group, thereby reducing the variability in the intelligence test by 50 per cent. Then, on substituting $\frac{s_1}{\sigma_1} = 0{\cdot}5$ and $r_{12} = 0{\cdot}6$ in the above formula, it can be seen that the resulting correlation between the two tests for the selected boys is reduced to 0·35.

§ 15. ASSOCIATION AND CONTINGENCY

Association.—Sometimes an attribute A is not quantitatively graded but is simply noted as being present or absent for each individual in the group. This is called by logicians, division by dichotomy. We may desire to know whether there is any sort of relationship between two such attributes A and B or whether they are quite unrelated.

	Typhoid	Not Typhoid	Total
Inoculated . .	56[a]	6,759[b]	6,815
Not Inoculated .	279[c]	11,396[d]	11,668
Total . .	328	18,155	18,483

[1] K. Pearson, "On the Influence of Natural Selection on the Variability and Correlation of Organs", Trans. of Roy. Soc. Series A, vol. 200, 1903.

Yule's *Coefficient of Association*, Q, gives a rough indication of the degree of association where

a is the number of cases in which both attributes are present,
b the number in which the first attribute is present and the second absent,
c the number in which the second is present and the first absent,
d the number in which both attributes are absent.

$$Q = \frac{ad - bc}{ad + bc}.$$

The standard error of Q is

$$\sigma_Q = \frac{1 - Q^2}{2} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}.$$

Even with attributes which are quantitatively graded, a measure of Q may be obtained by noting for each individual whether each attribute is above or below the average. In this way a rough estimate of the degree of correlation to be expected may be obtained, but Q yields unduly high values when compared with the product moment coefficient of correlation.

Another measure which is preferable as it yields values more in accordance with the product moment coefficient is Yule's *coefficient of colligation*, ω.

$$\omega = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}.$$

The standard error of ω is

$$\sigma_\omega = \frac{1 - \omega^2}{4} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}.$$

Contingency.—It is often necessary to test the agreement between observation and hypothesis. In the simplest case of a fourfold table, as in the above table which is taken from Greenwood and Yule's data for Typhoid, the probability that a value will fall in the first column is $\frac{328}{18{,}483}$, and the probability that a value will fall in the first row is $\frac{6815}{18{,}483}$. If these

two events were independent, the probability for their combined occurrence is the product of the two probabilities or $\frac{328 \times 6815}{(18{,}483)^2}$. Out of the 18,483 values we should, therefore, expect $18{,}483 \times \frac{328 \times 6815}{(18{,}483)^2}$ to fall in the first quadrant. The difference, $56 - \frac{328 \times 6815}{18{,}483}$, is a measure of the departure of the two attributes, inoculation and attack of typhoid, from complete independence. Similarly for the other three quadrants. Now calculate $\frac{(f - f_e)^2}{f_e}$ for each quadrant, where f is the observed frequency, and f_e the expected frequency on the hypothesis of independence. The sum of these values is called the *Chi Square Function* or χ^2.

Since $\chi^2 = 56{\cdot}234$ it can be seen from Fisher's Table [1] of χ^2 that the observations are clearly opposed to the hypothesis of independence. In actual practice χ^2 may, for fourfold tables, be obtained directly from the formula

$$\chi^2 = \frac{(ad - bc)^2(a + b + c + d)}{(a + b)(c + d)(a + c)(b + d)}.$$

It is instructive to study the significance of the values in the above table as determined by Q and σ_Q, by ω and σ_ω, and by χ^2 and Fisher's Table.

Tests of goodness of fit, using the χ^2 function, are somewhat beyond the scope of this book.[2]

§ 16. THE TETRAD DIFFERENCE CRITERION

Spearman applied tests of different mental abilities to a number of individuals and measured the inter-correlations. He then made the momentous discovery that if any four of these tests be considered, their correlations tended towards a peculiar

[1] R. A. Fisher, *op. cit.*, p. 104. Approximate charts also appear in H. Banister, *Elementary Applications of Statistical Method*, 1929, p. 42, and in S. Dawson, *Computation of Statistics*, 1933, p. 116.

[2] R. A. Fisher, *op. cit.* Chapter 4.

arrangement which could be expressed in a definite mathematical formula, namely:

$$r_{12}r_{34} - r_{13}r_{24} = 0.$$

This is called the tetrad equation and the value of $r_{12}r_{34} - r_{13}r_{24}$ is the *tetrad difference*.[1]

The next step was mathematical, namely, to demonstrate that whenever the tetrad equation holds, within the limits of fluctuations of sampling, and only when it does so, then every individual measurement of each ability will be divisible into two factors, namely, the general factor "g" and a specific factor "s". The "g", although varying freely from individual to individual, yet remains the same for any one individual in respect of all the correlated abilities. The "s" not only varies from individual to individual but also for any one individual from each ability to another.

The precise mathematical expression[2] of the divisibility into two factors is given by the equation:

$$m_{ax} = r_{ag}g_x + r_{as}s_{ax}$$

where m_{ax} is the measurement obtained for an individual x in the test of variable a; g_x is the individual's amount of the "g" factor; and s_{ax} is his amount of the factor specific to the test of variable a. Kelley[3] has expressed Spearman's proposition in the form: Four variables may be thought of as due to one general factor plus four specific factors when $r_{12}r_{34} = r_{13}r_{24} = r_{14}r_{23}$.

Kelley has also introduced the following convenient notation:

$$t_{1234} = r_{12}r_{34} - r_{13}r_{24} = 0,$$
$$t_{1342} = r_{13}r_{24} - r_{14}r_{23} = 0,$$
$$t_{1423} = r_{14}r_{23} - r_{12}r_{34} = 0.$$

[1] C. Spearman, *The Abilities of Man*, Appendix, p. iii, 1927.

[2] C. Spearman, *ibid.* p. xiv.

[3] T. L. Kelley, *Crossroads in the Mind of Man*, 1928, p. 46. Cf. also K. J. Holzinger, *Statistical Résumé of the Spearman Two-Factor Theory*, 1930, p. 3.

Of these three conditions only two are independent, since $t_{1234}+t_{1342}+t_{1423}=0$.

It is obvious that the other three possible tetrads, namely, t_{1243}, t_{1324}, and t_{1432}, are merely the first three with the signs reversed. When Spearman's tetrad difference criterion is applied in actual practice, it is of course never expected that a tetrad difference should be exactly zero. What is done is to compare its value with its standard error in order to see if its value is significant.

The standard error of t_{1234} may be written, if n is the number of cases in the sample

$$\sigma_{t_{1234}}=\frac{1}{\sqrt{n}}\{r^2_{12}+r^2_{13}+r^2_{24}+r^2_{34}-2(r_{12}r_{13}r_{23}+r_{12}r_{14}r_{24}$$
$$+r_{13}r_{14}r_{34}+r_{23}r_{24}r_{34})+4r_{12}r_{13}r_{24}r_{34}\}^{\frac{1}{2}}.$$

Other formulae which may be applied in suitable cases are given on pp. x, xi in the Appendix to *The Abilities of Man*.

CHAPTER 24

PSYCHOPHYSICAL METHODS

It is common knowledge that when a pound weight is held in the hand of a blindfolded person he might detect it to be lighter than another which weighed 18 ounces, although he would not be able to detect this addition of two ounces to a weight of ten pounds. Clearly then it is not the difference between the weights to be compared which is the crucial factor. What then can it be? It is rather remarkable that it is only about a hundred years since the underlying principle was discovered.

WEBER'S LAW

E. H. Weber (1795–1878), professor of anatomy at Leipzig, found that a weight of 30 ounces could just barely be distinguished as lighter than a weight of 31 ounces, but that a weight of 15 ounces could just barely be distinguished as lighter than a weight of $15\frac{1}{2}$ ounces. From his experiments Weber concluded that the just noticeable difference between two stimuli R and $R+dR$ depends not on the actual difference but on the ratio of that difference to the magnitude of R. In other words, Weber's Law may be expressed by the equation

$$\frac{dR}{R}=C,$$

where C is a constant. Obviously the law does not hold when the stimuli are very large or very small. For lifted weights the value of C is about $\frac{1}{30}$, for pressures $\frac{1}{20}$, and for estimating the lengths of lines by eye $\frac{1}{100}$. Weber's Law holds approximately within wide limits throughout the whole field of sensations. An apparent exception is that a person who can just distinguish

the tones of 200 and $200\frac{1}{2}$ vibrations from one another will just be able to distinguish the tones of 400 and $400\frac{1}{2}$. Why is this?[1]

Different interpretations have been given to the term "just noticeable difference". Some take it to mean a difference which in a long series of trials is as often noticeable as unnoticeable, that is, in 50 per cent of the trials it would be noticeable. Others take it to be that which in a long series of trials is judged correctly 75 per cent, for this lies half-way between 50 per cent correct which would result from mere chance, and 100 per cent correct which denotes certainty. Still others adopt 80 per cent correct. In the following paragraphs the 50 per cent basis has been adopted except in the Method of Serial Groups.

Within certain limits Weber's Law holds not only for just noticeable but also for easily noticeable differences in stimulus magnitude. Thus, with a light of 10 candle-power and another of 1000 candle-power, when S has to adjust a third variable light so that the intensity of its brightness appears midway between the two, the candle-power will approximate, not to 505 but to 100, *i.e.* 10 : 100 : : 100 : 1000.

FECHNER'S LAW

G. T. Fechner (1801–1887), for some years professor of physics at Leipzig, may be regarded as the founder of psychophysics. In his *Elemente der Psychophysik* published in 1860 he attempted to develop a science of the functional relation between body and mind. Fechner made three assumptions which have been subjects of controversy ever since. Incidentally he had satisfied himself that Weber's results were valid between wide limits. Fechner assumed (1) that a sensation is a measurable magnitude, the sum of a certain number of sensation units, (2) that the just noticeable difference of sensation may be treated as the difference between two sensations, and (3) that all just noticeable differences of sensation are equal, from whatever part of the stimulus scale they may be taken.

By far the most comprehensive discussion of Fechner's Law is to be found in Titchener's classical treatise[2] which is essential

[1] C. S. Myers, *Textbook of Experimental Psychology*, 1911, p. 246.

[2] E. B. Titchener, *Experimental Psychology*, vol. 2, Part 1, Introduction, and vol. 2, Part 2, Introduction, 1905.

for all serious students of the topic. It is true that in recent times Fechner's Law is not so much a question of the day as it used to be. In the first place, this is due to the shift of attention from the more academical to the more practical problems of our science. Yet there are signs now that such problems, not only in psychology but also in other sciences, may very profitably be studied in the light of Fechner's Law. In the second place, it was apparently easy in the view of most psychologists to refute Fechner's assumptions on introspective grounds. At the same time, the law itself as distinct from the assumptions on which it was based, does not fare at all badly at the bar of introspection. Hence there arises a dilemma.

William James's well-known criticism of the first assumption: "To introspection, our feeling of pink is surely not a portion of our feeling of scarlet",[1] agrees well with the verdicts of Stumpf, Münsterberg, Ebbinghaus, Sully, Külpe and many others.[2]

Myers, for example, criticises the second assumption and holds that the just noticeable difference is an experience of difference and not a difference between two sensations. "The experience that two stimuli are identical or different is as strictly *sui generis* as the experience of a stimulus itself."[3]

The third assumption has likewise given rise to endless arguments, but was admitted with reservations by no less an authority than Titchener. It is beyond the scope of this book to discuss such criticisms in detail or to consider how they affect the validity of Fechner's Law.

Many of the objections to the law may be avoided by substituting for Fechner's "sensation" the idea of "sense-distance" which was introduced by Delboeuf[4] in 1873. Thus, given two grays, A and B, differing in brightness, it is possible to find another, C, so that the interval or sense-distance AC is equal to the sense-distance CB. In this way a series of grays may be arranged in order of brightness and the stimulus-values corresponding to each may be determined.

[1] W. James, *Principles*, 1, p. 546.
[2] E. B. Titchener, *Exp. Psychol.* vol. 2, Part 2, lii.
[3] C. S. Myers, *Exp. Psychol.* p. 251.
[4] J. R. L. Delboeuf, *Étude psychophysique*, 1873; *Revue philos.*, 1878.

Critics of psychical measurement sometimes make unreasonable demands. As Ferguson[1] has pointed out, there are many measurable physical magnitudes, such as temperature, which cannot be conceived as being built up of units spatially juxtaposed. Yet measurements of temperature are daily made.

On these assumptions Fechner derived his Fundamental Formula

$$dS = C\frac{dR}{R},$$

where dS is the small increment of sensation due to the increment of the stimulus R by dR. On integrating,

$$S = c \log_e R + C,$$

where e is the base of natural logarithms, and C is a constant. If r is the value of R for which the sensation is just below the threshold of consciousness,

$$0 = c \log_e r + C.$$

Hence
$$S = c \log_e \frac{R}{r}.$$

On changing to common logarithms,

$$S = k \log \frac{R}{r},$$

where k is a constant. This is Fechner's Metric Formula. If r be made the unit of R in accordance with Fechner's assumption

$$S = k \log R,$$

which is the usual mathematical expression of Fechner's Law.

Fechner also played the chief part in devising the psychophysical methods shortly to be described, a part which stamps him as a rare genius. But it is also well to emphasise that only a genius could have formulated the law which bears his name even if the insecure nature of his assumptions be admitted. It has given rise to controversies which have already lasted over

[1] A. Ferguson, "Quantitative Estimates of Sensory Events", *Nature*, 130, September 3, 1932.

seventy years, and although his law is not so prominent in modern psychology, yet the principle can be applied in practice over a wide range of experiences.

The accompanying figure illustrates how a rough attempt may be made to ascertain the validity of Fechner's Law, and incidentally to make its meaning clear.

Along the vertical axis, let OR_1 represent a weight of 100 grams. Along the horizontal axis, let OS_1 be any length to represent the strength or intensity of the sensation in lifting it. Now slowly add fine sand to another 100 gram weight until the subject judges it to be just noticeably heavier. Then, according to Fechner, we may take an arbitrary distance $S_1\ S_2$ to represent the unit of sensation, namely, the just noticeable distance. Let OR_2 represent this increased weight. This, in turn, is taken as the standard, and the comparison weight is increased until it is just noticeably heavier, say OR_3, corresponding to another unit of sensation $S_2\ S_3$. Similarly for OR_4, etc. According to Fechner's Law the resulting points P_1, P_2, P_3, . . . should lie on the curve $S = k \log R$.

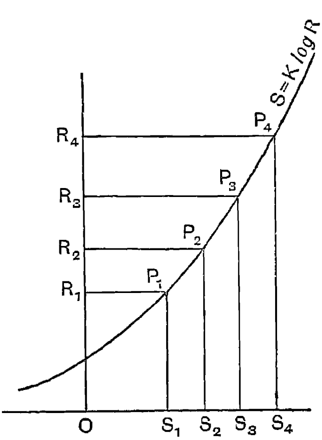

FIG. 33.

Thus, if $OR_1 = 100$ grams and $OR_2 = (100 + x)$ grams $= 100(1 + \frac{x}{100})$ grams. Then, by Fechner's Law,

$$OR_3 = 100(1 + \frac{x}{100})^2$$

and $$OR_4 = 100(1 + \frac{x}{100})^3, \text{ etc.}$$

For example, if OR_2 is 110 grams, so that $x = 10$, then OR_3 should be about 121 grams, and OR_4 about 133 grams.

THE PSYCHOPHYSICAL METHODS

Quantitative work in psychology often demands methods of measurement which were very largely developed by the genius of Fechner, namely, the psychophysical methods.[1] They are: (1) The Method of Mean Error. (2) The Method of Limits. (3) The Constant Method.

THE METHOD OF MEAN ERROR.—A standard stimulus is presented and S adjusts a variable stimulus until it appears equal to the standard. Cattell and Fullerton[2] criticised the method as the time taken by S to make the adjustment is not under control and, indeed, the degree of control must be considered in assessing any psychophysical method. The nature of

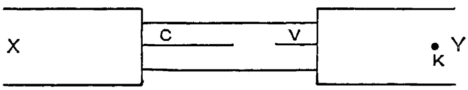

FIG. 34.

the possible errors may be understood by working out a numerical example.

C is the standard black line, 50 millimetres long, drawn on white celluloid and V is the variable line whose length may be altered by moving the frame Y by means of the knob K. The gap between C and V is 20 millimetres. Starting from a position where V is appreciably longer than C, S adjusts V until it appears equal to C. Then starting from a position where V is appreciably shorter than C he adjusts it equal again. This procedure is repeated until ten adjustments are made. If they were 49, 48, 51, 48·5, 48, 50·5, 49·5, 48, 49, 48, the mean or $r_{m_1}=49$. Under these conditions, then, 49 millimetres is the most probable value of what S regards as the equivalent of C. It is a value which tends to be free of accidental errors for these will in the long run cancel out, some being on the plus and some on the minus side. By summing up the differences regardless of sign between his estimates and his mean estimate we

[1] The advanced student will find an exhaustive account of these methods in E. B. Titchener's *Experimental Psychology*, vol. 2, Part 2, chap. 2, The Metric Methods: 1905.

[2] G. S. Fullerton and J. M'K. Cattell, *On the Perception of Small Differences*, 1892.

know how accurately S performs the task. These are: 0, 1, 2, 0·5, 1, 1·5, 0·5, 1, 0, 1, so that the mean variable error or $e_{m_1} = 0{\cdot}85$.

Now the apparatus is inverted so that V is on the left. If the corresponding adjustments are 52, 51·5, 49·5, 52, 52, 49·5, 50, 50·5, 51·5, 52·5, then the mean estimate or $r_{m_2} = 51{\cdot}2$ and the mean variable error or $e_{m_2} = 1{\cdot}06$.

But even if accidental errors have been eliminated, the estimates will almost certainly be affected by constant errors. In the first case, the *crude constant error* or $c_1 = r_{m_1} - 50 = -1$. In the second case, the *crude constant error* or $c_2 = r_{m_2} - 50 = 1{\cdot}2$. It is usual to regard the error as positive or negative according as the mean estimate is greater or less than the standard.

Now a part of this error is due to the spatial position of the two stimuli and is called the *space error* q, and it was in order to eliminate this part that the experiment was repeated with the standard on the right. But if q was the only error to be considered, $\frac{r_{m_1} + r_{m_2}}{2}$ should be 50. As this is not the case, the constant errors must have a second component s which remains unchanged throughout both procedures, that is:

$$c_1 = -q + s \qquad\qquad r_{m_1} - 50 = -q + s$$
$$c_2 = +q + s \qquad\qquad r_{m_2} - 50 = +q + s.$$

Since $c_1 = -1$ and $c_2 = 1{\cdot}2$ therefore $q = 1{\cdot}1$ and $s = 0{\cdot}1$.

Such an experiment would only serve for preliminary practice. 20 further adjustments with V on the left followed by 20 with V on the right should be taken. When this has been repeated on another day it can be ascertained whether the assumption of a constant error is justified and not due to chance owing to the number of adjustments being too small.

If justified, c_1 will have the same sign each day. Similarly for c_2. Further, the probable error of the average r_m is $\frac{0{\cdot}8453}{\sqrt{n}} \cdot e_m$ (cf. Chapter 23),[1] and if the error is considerably larger than this, it may be taken to represent a true constant error.

[1] For standard error $= 1{\cdot}2533 \times$ mean error
and probable error $= 0{\cdot}6745 \times$ standard error.

THE METHOD OF LIMITS.—The method can be employed to determine either the absolute threshold or the differential threshold. An example of the first would be to determine the smallest intensity of a stimulus to produce a sensation. An example of the second would be to find the smallest difference in intensity between two stimuli which can be apprehended. Two examples are worked out as illustrations:

(*a*) *The Absolute Threshold for Colour Saturation.*—The figure refers to the experiment described in Chapter 13, Section 1. The two interlocking discs are red, and gray of an equal brightness. The problem is to find how much red must be exposed before S can detect it. As the red sector is decreased, the red colour becomes less and less saturated and it becomes increasingly difficult to judge whether there is a trace of red present. At some stage S can see no trace of it. E starts from a point, say 11°, where S definitely reports "red" and exhibits in succession 10°, 9°, 8°, . . . of red and in each case S says either "yes" if he can see a trace of red, or "no" if he cannot, or "undecided" if he cannot come to a decision. Then E repeats the procedure but proceeding in the opposite direction from a point where S definitely answers "no", say 3°.

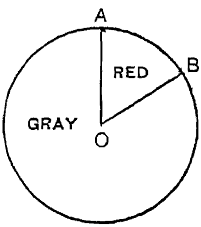

FIG. 35.

		b				*a*				
	3°	4°	5°	6°	7°	8°	9°	10°	11°	
	N	N ↓	U	N	Y	U ↓	Y	Y	Y	←
→	N	N	N ↑	U	U	Y	U ↑	Y	Y	
			d				*c*			

The upper threshold or the point where the region of uncertainty may be said to begin is $\frac{a+c}{2}$ or $\frac{8{\cdot}5^\circ+9{\cdot}5^\circ}{2}$ or 9°. Similarly the lower threshold is $\frac{b+d}{2}$ or 5°. The average threshold might then be regarded as $\frac{9^\circ+5^\circ}{2}$ or 7°. The whole procedure must be repeated several times and the averages considered. If S gives no "undecided" answers the procedure is obviously simplified.

(*b*) *The Differential Threshold for Lifted Weights.*—The standard weight C of 100 grams is placed to the right of the weight 114 grams. S lifts the standard and then lifts the other in the same way and judges whether it is heavier, lighter or indistinguishable in weight from the standard. S always judges in terms of the weight lifted second. Similarly 112 grams, 110 grams, etc. are compared with the standard until S cannot distinguish between them. Then E starts with a weight equal to the standard and increases the variable until S consistently answers "heavier".

86	88	90	92	94	96	98	100	102	104	106	108	110	112	114
											$u\uparrow_a$	h	h	$h\leftarrow$
							$\rightarrow u$	u	u	$u\uparrow_c$	h			
$\rightarrow l$	$l\uparrow_b$	u												
		$l\uparrow_d$	u	u	u	u	$u\leftarrow$							

Thus Tu_1, the upper threshold, is $\frac{109+107}{2}$ grams or 108 grams.

Similarly, Tl_1, the lower threshold, is $\frac{89+91}{2}$ grams or 90 grams.

The above procedure should be repeated at least ten times. Now the value of the threshold depends on the temporal and spatial arrangement of the standard and variable weights. The time error e_1 is positive or negative according as the effect of the time order is to increase or diminish the apparent value of the *first-presented* stimulus. The space error e_2 is positive or negative according as the effect of the space order is to increase or diminish the apparent value of the *left-hand* stimulus.[1] The more advanced student may therefore arrange the experiment as follows:

I.	Standard	presented	first	and to the	right.
II.	,,	,,	second	,,	,,
III.	,,	,,	first	,,	left.
IV.	,,	,,	second	,,	,,

From the resulting equations the time error and the space error may be calculated.[2].

[1] W. Brown and G. H. Thomson, *The Essentials of Mental Measurement*, 1921, p. 49. [2] *Ibid.*

THE METHOD OF SERIAL GROUPS.—As the variable is changed progressively in the method of limits it is not surprising that S's judgment is influenced by expectation. If attempts are made to eliminate this by the interpolation of "catch" stimuli in an irregular manner there is an uncertain disturbing effect on S, but there is one modification of the method of limits, namely, the method of serial groups, where the "catch" stimuli are systematically interpolated. It was developed independently by McDougall[1] and by Stratton,[2] and the method will now be illustrated with reference to:

The Absolute Spatial Threshold.—S rests his arm on the table with the palm downwards. An ink-dot is placed on the arm about 4 in. above the wrist. E places the two points of an aesthesiometer, with equal and constant pressure, lengthwise of the arm on either side of the ink-dot. S, with closed eyes, judges whether he is touched by two points or one. After preliminary exploration with the points set at various distances, suppose it is found that S becomes certain in his judgments that there are two touches when the points are 50 millimetres apart. Then the distance is reduced by 5 millimetres and S is now stimulated twenty times either by the one point C, applied near one or other of the two spots to which the two points are applied, or else simultaneously by the two points A and B. If he correctly judges each of the ten two-points stimulations, the distance between the points is further reduced by 5 millimetres, and so on until S fails to get eight out of the ten correct. It is arbitrary to use the 80 per cent point rather than the 50 per cent point but it has the advantage of quickness in practice. The threshold is taken to be the point where 80 per cent of the judgments are correct. If this were 30 millimetres it is expedient to write the judgments for two-point stimulation on the top line and those for one-point underneath:

30mm.	T_1	T_2	U_5	T_7	T_8	O_{11}	T_{12}	T_{13}	T_{17}	T_{19}
	O_3	O_4	U_6	O_9	O_{10}	O_{14}	O_{15}	O_{16}	O_{18}	O_{20}

[1] Rep. Cambridge Anthrop. Expedition to Torres Straits. II., 1903, p. 190.
[2] G. M. Stratton, "The Method of Serial Groups", *Psychol. Rev.* vol. 9, 1902, p. 444.

The spatial threshold[1] varies widely according to the region tested. Approximate values are as follows: tip of tongue 1 mm., tip of finger 2 mm., back of hand 30 mm., spine 54 mm.

The Differential Threshold for Lifted Weights may also be found by the method of serial groups. Thus if preliminary testing made it expedient to start with the comparison of 112 grams with the standard of 100 grams, then ten such comparisons would be interspersed with comparisons of 100 grams with another 100 grams. This procedure would be repeated with a comparison of 110 grams and so on until the threshold was found.

THE CONSTANT METHOD.—This is also known as the method of "right and wrong cases". It is the most exact of the psychophysical methods and like the method of limits it can be employed for determining either an absolute or differential threshold. By preliminary experiments the different values of the variable stimulus to be employed are fixed at the start, and are presented to S a large number of times in irregular order, or in a prearranged order which is unknown to S. The range of stimuli presented must be great enough to ensure that S's judgments are always the same for the smallest stimulus presented, and likewise for the biggest. The method is best illustrated by application to actual examples:

(1) *The Absolute Threshold for Colour Saturation* (cf. Chapter 13).—The order of presentation of the stimuli will be clear from an inspection of the table. It is further obvious that the threshold lies between 4° and 5°. But it may be worth while to show how a more exact value may be obtained by means of Spearman's formula[2]

$$L = \frac{1}{2}(D_{s+1} + D_s) - \frac{i\Sigma g}{n},$$

where L is the threshold, D_{s+1} the smallest stimulus which yields no "gray" judgment, i is the interval between the stimuli (here 1°), g is the number of "red" judgments for all stimuli

[1] Cf. G. M. Whipple, *op. cit.* vol. 1, pp. 243-261.

[2] C. Spearman, "The Method of 'Right' and 'Wrong' Cases ('Constant Stimuli') without Gauss's Formulae", *Brit. Journ. of Psychol.* vol. 2, 1908, p. 235.

U

smaller than D_{s+1}, and n is the number of judgments for each stimulus (here 10):

2°	3°	4°	5°	6°
G 1	G 3	R 5	R 2	R 4
G 2	G 4	G 1	R 3	R 5
G 3	G 5	G 2	R 4	R 1
G 4	G 1	G 3	G 5	R 2
G 5	G 2	G 4	G 1	R 3
G 1	G 3	G 5	G 2	R 4
G 2	R 4	G 1	G 3	R 5
G 3	G 5	G 2	R 4	R 1
G 4	G 1	G 3	R 5	R 2
G 5	G 2	G 4	R 1	R 3
G = 10	9	9	4	0
R = 0	1	1	6	10

$$L=\frac{6+5}{2}-\frac{1(1+1+6)}{10}=4{\cdot}7°$$

The table is completed in rows according to the order shown. Thus 2° of red is first shown, then 5°, 3°, 6°, 4°.

A measure of the average deviation may be obtained by considering the 30 judgments under 3°, 4° and 5° which cover the region of uncertainty, and regarding a "red" judgment smaller than the threshold as an "error", and a "gray" judgment larger than the threshold also an "error", then

$$\begin{aligned}30\text{ A.D.} &= 1\times 1{\cdot}7° + 1\times 0{\cdot}7° + 4\times 0{\cdot}3°\\ \text{A.D.} &= 0{\cdot}12°\end{aligned}$$

It will be noticed that "undecided" answers were not given in this example, but the constant method will now be illustrated in the case of an example where "equal" or "undecided" answers are obtained.

(2) *The Differential Thresholds* (Example: Lifted Weights).—Assuming that the range of the variable extends from 110 to 88 grams and that each variable has been compared with the standard of 100 grams ten times, let the judgments be distributed as in the following table:

X_m	g_m	u_m	l_m	$g_m(1-g_m)$	$u_m(1-u_m)$	$g_m(1-l_m)$	$g_m+l_m-(g_m-l_m)^2$	m
110	1·0	0	0	0	0	0	1·0 – 1·0	11
108	0·9	0·1	0	0·09	0·09	0	0·9 – 0·81	10
106	0·9	0·1	0	0·09	0·09	0	0·9 – 0·81	9
104	0·7	0·3	0	0·21	0·21	0	0·7 – 0·49	8
102	0·6	0·3	0·1	0·24	0·21	0·09	0·7 – 0·25	7
100	0·5	0·4	0·1	0·25	0·24	0·09	0·6 – 0·16	6
98	0·1	0·4	0·5	0·09	0·24	0·25	0·6 – 0·16	5
96	0	0·3	0·7	0	0·21	0 21	0·7 – 0·49	4
94	0	0·2	0·8	0	0·16	0·16	0·8 – 0·64	3
92	0	0·2	0·8	0	0·16	0·16	0·8 – 0·64	2
90	0	0·1	0·9	0	0·09	0·09	0·9 – 0·81	1
88	0	0	1·0	0	0	0	0	0
..	4·7	2·4	4·9	0·97	1·70	1·05	2·34	..

Thus the weight 98 grams has been judged heavier once, lighter 5 times, and undecided 4 times. Therefore $g_m + u_m + 1_m = 1$. The upper threshold or limen L_{up} may be found by the use of Spearman's formula

$$L_{up} = \frac{1}{2}(D_{s+1} + D_s) - \frac{i\Sigma g}{n}$$

$$= \frac{1}{2}(110 + 108) - \frac{2(9+9+7+6+5+1)}{10} = 101{\cdot}6 \text{ grams.}$$

Similarly the lower threshold or L_{lo} is

$$L_{lo} = \frac{1}{2}({}_{t+1}D + {}_tD) + \frac{i\Sigma l}{n}$$

$$= \frac{1}{2}(88 + 90) + \frac{2(9+8+8+7+5+1+1)}{10} = 96{\cdot}8 \text{ grams.}$$

Very similar is the procedure when using Wirth's formula[1]

$$L_{up} = X_m - i(\Sigma g_m - \frac{1}{2})$$

$$= 110 - 2(4{\cdot}7 - 0{\cdot}5) = 101{\cdot}6 \text{ grams.}$$

$$L_{lo} = X_o + i(\Sigma l_m - \frac{1}{2})$$

$$= 88 + 2(4{\cdot}9 - 0{\cdot}5) = 96{\cdot}8 \text{ grams.}$$

[1] W. Wirth, *Spezielle psychophysische Massmethoden*, 1927.

Thus the average differential threshold or DL is given by

$$\mathrm{DL}=\frac{1}{2}(\mathrm{L}_{up}-\mathrm{L}_{lo})=\frac{1}{2}(101{\cdot}6-96{\cdot}8)=2{\cdot}4 \text{ grams}$$

$\mathrm{L}_{up}-\mathrm{L}_{lo}$ is Müller's measure[1] of the "ideal region of u-judgments" and is equal to $\frac{i\Sigma u}{n}$.

The subjective equivalent of the standard weight of 100 grams is

$$\mathrm{E}=\frac{1}{2}(\mathrm{L}_{up}+\mathrm{L}_{lo})=\frac{1}{2}(101{\cdot}6+96{\cdot}8)=99{\cdot}2 \text{ grams.}$$

Wirth's formulae for the probable errors of the above measures are:

$$\mathrm{PE}_{\mathrm{L}up}=0{\cdot}675\frac{i}{\sqrt{n}}\sqrt{\Sigma g_m(1-g_m)},$$

$$\mathrm{PE}_{\mathrm{L}lo}=0{\cdot}675\frac{i}{\sqrt{n}}\sqrt{\Sigma l_m(1-l_m)},$$

$$\mathrm{PE}_{\mathrm{DL}}=0{\cdot}675\frac{i}{2\sqrt{n}}\sqrt{\Sigma u_m(1-u_m)},$$

$$\mathrm{PE}_{\mathrm{E}}=0{\cdot}675\frac{i}{2\sqrt{n}}\sqrt{\Sigma[(g_m+l_m)-(g_m-l_m)^2]}.$$

Hence

$$\mathrm{L}_{up}=101{\cdot}6\pm0{\cdot}675\frac{2}{\sqrt{10}}\sqrt{0{\cdot}97},$$

$$\mathrm{L}_{lo}=96{\cdot}8\pm0{\cdot}675\frac{2}{\sqrt{10}}\sqrt{1{\cdot}05},$$

$$\mathrm{DL}=2{\cdot}4\pm0{\cdot}675\frac{2}{2\sqrt{10}}\sqrt{1{\cdot}70},$$

$$\mathrm{E}=99{\cdot}2\pm0{\cdot}675\frac{2}{2\sqrt{10}}\sqrt{2{\cdot}34}.$$

(3) *The Differential Threshold when the steps are unequal* (Example: Lifted Weights).—This method also was devised by

[1] G. E. Müller, *Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik*, 1904, p. 148.

Spearman.[1] Here only the approximate form of the method is given. For the more exact procedure Professor Spearman's paper should be consulted.

Let the standard weight S be 100 grams and the comparative weight C vary by unequal steps from 82 to 118 grams. Let C − S = D and the distribution of judgments be as in the table:

C	D	g	u	l	g^1	l^1	c	p	q	cp +	cp −	cq +	cq −
82	− 18	0	0	10	0	10							
86	− 14	0	1	9	½	9½	− 16	0	1				− 16
89	− 11	1	1	8	1½	8½	− 12½	1	1		− 12½		− 12½
91	− 9	1	2	7	2	8	− 10	0	1		0		− 10
96	− 4	2	4	4	4	6	− 6½	1	3		− 6½		− 19½
100	0	5	3	2	6½	3½	− 2	3	2		− 6		− 4
103	+ 3	6	2	2	7	3	− 1½	1	0	1½		0	
107	+ 7	7	2	1	8	2	− 5	1	1	5		5	
110	+ 10	8	1	1	8½	1½	− 8½	1	0	8½		0	
114	+ 14	9	1	0	9½	½	− 12	1	1	12		12	
118	+ 18	10	0	0	10	0	− 16	1	0	16		0	
..	..	..	..	..	..	..	..	..	..	+ 18		− 45	

Now the upper threshold may be regarded as corresponding to the exact value of D at which the subject should just balance between judging C to be equal to or greater than S. Thus when D is +7, 70 per cent of the thresholds were less than +7, and when D is +10, 80 per cent were less than +10. Therefore 10 per cent must occur between +7 and +10. The column headed p thus gives the percentage of thresholds between successive limits. In the same way q is the corresponding percentage for the lower threshold. The c column gives the middle value between two successive Ds. Hence

$$L_{up} = \frac{\Sigma cp}{n} = \frac{18}{10} = 1{\cdot}8 \text{ grams},$$

$$L_{lo} = \frac{\Sigma cq}{n} = \frac{-45}{10} = -4{\cdot}5 \text{ grams},$$

[1] C. Spearman, "The Method of 'Right' and 'Wrong' Cases ('Constant Stimuli') without Gauss's Formulae", *Brit. Journ. of Psychol.* vol. 2, 1908.

$$\text{DL}=\frac{\text{L}_{up}+\text{L}_{lo}}{2}=\frac{1\cdot 8+4\cdot 5}{2}=3\cdot 15 \text{ grams,}$$

$$\text{E}=101\cdot 8-3\cdot 15=95\cdot 5+3\cdot 15=98\cdot 65 \text{ grams.}$$

This result may be checked by calculating $\frac{\Sigma cp^1}{n}=\frac{\Sigma cq^1}{n}$, where p^1 and q^1 are obtained from the g^1 and l^1 columns.

References

1. G. T. Fechner, *Elemente der Psychophysik*, 1860, 2nd ed., 1889.
2. J. R. L. Delbœuf, *Éléments de psychophysique, générale et spéciale*, 1883.
3. G. E. Müller, *Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik*, 1904.
4. W. Wundt, *Grundzüge der physiologischen Psychologie*, 1, chap. 9, 1908.
5. E. B. Titchener, *Experimental Psychology*, 2, Parts 1 and 2, 1905.
6. C. Spearman, "The Method of 'Right' and 'Wrong' Cases ('Constant Stimuli') without Gauss's Formulae", *Brit. Journ. of Psychol.* vol. 2, 1908.
7. F. M. Urban, "Die psychophysischen Massmethoden als Grundlage empirischer Messungen", *Arch. f. d. ges. Psychol.* 15 and 16, 1909.
8. W. Brown and G. H. Thomson, *The Essentials of Mental Measurement*, 1921.
9. W. Wirth, *Spezielle psychophysische Massmethoden in Abderhaldens Handbuch der biologischen Arbeitsmethoden*, 1927.

NAME INDEX

SUBJECT INDEX

THE END

Printed in Great Britain by R. & R. Clark, Limited, *Edinburgh.*